

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 04:22:47 ; Search time 4479.5 Seconds  
(without alignments)  
16284.480 Million cell updates/sec

Title: US-09-912-559-1  
Perfect score: 1683  
Sequence: 1 agtggccagagatgtctga.....aaagtgaagtgccttctaa 1683

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_ov.\*  
22: em\_pat.\*  
23: em\_ph.\*  
24: em\_pl.\*  
25: em\_ro.\*  
26: em\_sts.\*  
27: em\_un.\*  
28: em\_un.\*  
29: em\_un.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1683	100.0	1683	6	AX383954 Sequence
2	1683	100.0	1683	6	BD177602 Detection
3	1683	100.0	2408	9	S83182 Hyaluronan-
4	1683	100.0	3008	6	AX409021 Sequence
5	1683	100.0	3008	9	HUMHGFA1 Human mRNA
6	1681.4	99.9	2251	9	BC031412 Homo sapi
7	1679.8	99.8	1683	6	AX383955 Sequence
8	1679.8	99.8	1683	6	BD177603 Detection
9	1144.2	68.0	2190	10	AK128915 Mus muscu
10	1050.2	62.4	2119	10	BC031775 Mus muscu
11	261.2	15.5	188349	9	AL390197 Human DNA
12	258	15.3	157750	2	AC006097 Homo sapi
13	242.6	14.4	397	6	AX305372 Sequence
14	167.4	9.9	204214	2	AC115771 Mus muscu
15	161	9.6	224208	2	AC106236 Rattus no
16	161	9.6	24249	2	AC131859 Rattus no
17	133	7.9	2035	10	AF099017 Mus muscu
18	133	7.9	2063	10	AF224724 Mus muscu
19	131.4	7.8	2067	10	BC019376 Mus muscu
20	117.6	7.0	2033	6	E07591 DNA encodin
21	117.6	7.0	2033	6	I15469 Sequence 14
22	117.6	7.0	2033	6	I69315 Sequence 14
23	117.6	7.0	2036	6	AX333070 Sequence
24	117.6	7.0	2036	6	AX408965 Sequence
25	117.6	7.0	2036	6	AX408965 Sequence
26	100.6	6.0	2375	4	HUMHGFA1 Homo sapien
27	99.8	5.9	2228	4	BTUPA Porcine mRN
28	99.8	5.9	2355	4	BOVUKPA Bos taurus
29	96.2	5.7	970	6	E07615 DNA encodin
30	96.2	5.7	970	6	I15462 Sequence 3
31	96.2	5.7	970	6	I69308 Sequence 3
32	94.6	5.6	1454	10	RNUFAM R.norvegicu
33	94.6	5.6	2366	10	RNUPLAC R.norvegicu
34	93.2	5.5	2299	10	MMURRR Mouse mRN
35	90.6	5.4	2512	6	A19618 Mammalian c
36	90.6	5.4	2585	10	BC061565 Rattus no
37	89	5.3	2445	10	RATPATISS Rat tissue-
38	87.6	5.2	2245	4	VMPAA1A Desmodus ro
39	87.6	5.2	2252	6	A07242 Artificial
40	87.6	5.2	2252	6	A07250 Artificial
41	87.6	5.2	2266	4	VMPAA1S Desmodus ro
42	84.8	5.0	2130	4	VMPAB Desmodus ro
43	84.8	5.0	2137	6	A07245 Artificial
44	81.2	4.8	1620	6	I08184 Sequence 4
45	81.2	4.8	1725	4	VMP1PA Vampire bat

ALIGNMENTS

RESULT 1  
AX383954  
LOCUS AX383954 1683 bp DNA linear PAT 19-MAR-2002  
DEFINITION Sequence 1 from Patent EP1182258.  
ACCESSION AX383954  
VERSION AX383954.1 GI:19577525  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1  
AUTHORS Roemisch, J., Stoeck, H.A., Feusner, A., Lang, W., Weimer, T.,  
Becker, M., Nerlich, C. and Muth-Naumann, G.  
TITLE Mutants of the factor vii activating protease and methods for their

Pred. No. is the number of results predicted by chance to have a

detection		Patent: EP 1182258-A 1 27-FEB-2002;
JOURNAL		Aventis Behring GmbH (DE)
FEATURES		Location/Qualifiers
source		1..1683
		/organism="Homo sapiens"
		/mol_type="unassigned DNA"
		/db_xref="taxon:9606"
ORIGIN		
Query Match		100.0%; Score 1683; DB 6; Length 1683;
Best Local Similarity		100.0%; Pred. No. 0;
Matches 1683; Conservative		0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	ATGTTTGCAGGATGCTGATCCTCCATGTTTCGCTGTTAAATGGCTCTCGTGGGAAAGACA 60
Db	1	ATGTTTGCAGGATGCTGATCCTCCATGTTTCGCTGTTAAATGGCTCTCGTGGGAAAGACA 60
Qy	61	GCCTGTGGGTTCTCCCTGATGCTTTATTTGGAAGCCCTGGACCCAGACTGGACCCCTGAC 120
Db	61	GCCTGTGGGTTCTCCCTGATGCTTTATTTGGAAGCCCTGGACCCAGACTGGACCCCTGAC 120
Qy	121	CAGTATGATTACAGCTACGAGATTATATCAGGAAGAGACACACAGTAGCACATTACC 180
Db	121	CAGTATGATTACAGCTACGAGATTATATCAGGAAGAGACACACAGTAGCACATTACC 180
Qy	181	CATGCTGAGATCCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
Db	181	CATGCTGAGATCCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240
Qy	241	CCCTGTGACAGGTTGGGACTGCTCTCCATGGAGACACTTCACATGACGTGCGTG 300
Db	241	CCCTGTGACAGGTTGGGACTGCTCTCCATGGAGACACTTCACATGACGTGCGTG 300
Qy	301	GCTCTCTTCTCTGGGAATTAAGTGTGAGAAAGTSCAAAATACGTGCAAGGACAAACCCATGT 360
Db	301	GCTCTCTTCTCTGGGAATTAAGTGTGAGAAAGTSCAAAATACGTGCAAGGACAAACCCATGT 360
Qy	361	GGCGGGGCAATGCTCAATACCCAGACTCTCCCTACTACCGCTGTCTGTAACAC 420
Db	361	GGCGGGGCAATGCTCAATACCCAGACTCTCCCTACTACCGCTGTCTGTAACAC 420
Qy	421	CTTTACACAGGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGCGCAAAACCCCTGCCAG 480
Db	421	CTTTACACAGGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGCGCAAAACCCCTGCCAG 480
Qy	481	AATGGGGTACCTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACTGTGCTGCTCCGAC 540
Db	481	AATGGGGTACCTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACTGTGCTGCTCCGAC 540
Qy	541	CAGTTCAAGGGAAATTCGTGAAATAGGTTCTGTGACTGCTATGTTGGCGATGGCTAC 600
Db	541	CAGTTCAAGGGAAATTCGTGAAATAGGTTCTGTGACTGCTATGTTGGCGATGGCTAC 600
Qy	601	TCTTACCGAGGAAATGAATAGGACAGTCAACCCAGCATGCTGCTCTTACTGGAATCC 660
Db	601	TCTTACCGAGGAAATGAATAGGACAGTCAACCCAGCATGCTGCTCTTACTGGAATCC 660
Qy	661	CACTCTCTCTGAGGAGATTAACACATGTTTATGGAGATGCTGAAACCCATGGGATT 720
Db	661	CACTCTCTCTGAGGAGATTAACACATGTTTATGGAGATGCTGAAACCCATGGGATT 720
Qy	721	GGGGAACACAAATTTCTGCAAGAACCCAGATGCGGACGAAAGCCCTGGTGTATTTAA 780
Db	721	GGGGAACACAAATTTCTGCAAGAACCCAGATGCGGACGAAAGCCCTGGTGTATTTAA 780
Qy	781	GTTTACCAATACAGAGTGAATGGGAATCTGTGATGCTCAGCTGCTCAGCCGAGAC 840
Db	781	GTTTACCAATACAGAGTGAATGGGAATCTGTGATGCTCAGCTGCTCAGCCGAGAC 840
Qy	841	GTTGCTCTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGTTCCGGGGTTGACTCC 900
Db	841	GTTGCTCTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGTTCCGGGGTTGACTCC 900

Qy	901	TGTGGAAGACTGAGATACGAGAGGAGATCAAGAGATCTATGGAGCTTTAAGAGC 960
Db	901	TGTGGAAGACTGAGATACGAGAGGAGATCAAGAGATCTATGGAGCTTTAAGAGC 960
Qy	961	ACGGCGGCAAGCAACCCATGGCAGGCTCCCTCCAGTCCCTGCTGCTCTGACCATCTCC 1020
Db	961	ACGGCGGCAAGCAACCCATGGCAGGCTCCCTCCAGTCCCTGCTGCTCTGACCATCTCC 1020
Qy	1021	ATGCCCCAGGCGCACTTCTGTGTGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
Db	1021	ATGCCCCAGGCGCACTTCTGTGTGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080
Qy	1081	GCCCACTGACCCGACATAAAAAACCAACATCTAAAGTGGTGTCTAGGGGACGAGACCTG 1140
Db	1081	GCCCACTGACCCGACATAAAAAACCAACATCTAAAGTGGTGTCTAGGGGACGAGACCTG 1140
Qy	1141	ARGAAGAGAGATTTTCATGAGCAGAGCTTTAGGTTGGAGAGATATTTCAAGTACAGCCAC 1200
Db	1141	ARGAAGAGAGATTTTCATGAGCAGAGCTTTAGGTTGGAGAGATATTTCAAGTACAGCCAC 1200
Qy	1201	TACAATGAAAGAGATGAGATTTCCCAACAATGATTTGCAATCTCAAGTTAAAGCCAGTG 1260
Db	1201	TACAATGAAAGAGATGAGATTTCCCAACAATGATTTGCAATCTCAAGTTAAAGCCAGTG 1260
Qy	1261	GATGGTCACTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTTGCCTGATGGTCC 1320
Db	1261	GATGGTCACTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTTGCCTGATGGTCC 1320
Qy	1321	TTTCCCTCTGGGAGTCAGTGCACATCTCTGGCTGGGTGTTACAGAAAACAGAAAAGG 1380
Db	1321	TTTCCCTCTGGGAGTCAGTGCACATCTCTGGCTGGGTGTTACAGAAAACAGAAAAGG 1380
Qy	1381	TCCCGGAGCTCTGATGCAAGTCAAGCTGATTTGCCAACTTTGTGCAACTCCCGC 1440
Db	1381	TCCCGGAGCTCTGATGCAAGTCAAGCTGATTTGCCAACTTTGTGCAACTCCCGC 1440
Qy	1441	CAACTCTATGACCAATGATTTGATGACAGTATGATCTGTGCAAGAAATCTTCAGAAACCT 1500
Db	1441	CAACTCTATGACCAATGATTTGATGACAGTATGATCTGTGCAAGAAATCTTCAGAAACCT 1500
Qy	1501	GGGCAAGACACCTGCGAGGTTGATCTGTGAGGCCCCCTGACCTGTGAGAGGACGCGACC 1560
Db	1501	GGGCAAGACACCTGCGAGGTTGATCTGTGAGGCCCCCTGACCTGTGAGAGGACGCGACC 1560
Qy	1561	TACTAGCTCTATGGGATAGTGAAGTGGGCTGGAGTGTGGGAAGAGGCGAGGGGTCTAC 1620
Db	1561	TACTAGCTCTATGGGATAGTGAAGTGGGCTGGAGTGTGGGAAGAGGCGAGGGGTCTAC 1620
Qy	1621	ACCAAGTTACCAAAATTCCTGAATGGATCAAGCCACCATCAAAAAGTGAAGTGGCTTC 1680
Db	1621	ACCAAGTTACCAAAATTCCTGAATGGATCAAGCCACCATCAAAAAGTGAAGTGGCTTC 1680
Qy	1681	TAA 1683
Db	1681	TAA 1683

RESULT 2  
BD177602 1683 bp DNA linear PAT 16-APR-2003  
LOCUS  
DEFINITION Detection method with the use of factor VII activating protease  
mutant and specific antibody.  
ACCESSION BD177602  
VERSION BD177602.1 GI:30014864  
KEYWORDS JP 2002291486-A/1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1683)  
Roemisch, J., Stoeck, H.A., Feussner, A., Lang, W., Weimer, T.,  
Becker, M., Nerlich, C. and Naumann, G.M.



QY 1681 TAA 1683  
Db 1681 TAA 1683

RESULT 3  
S83182 2408 bp mRNA linear PRI 11-FEB-1997  
DEFINITION hyaluronan-binding protein=hepatocyte growth factor activator  
S83182 homolog [human, plasma, mRNA, 2408 nt].  
ACCESSION S83182  
VERSION S83182.1 GI:1836158  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1. (bases 1 to 2408)  
Choi-Miura,N.H., Tobe,T., Sumiya,J., Nakano,Y., Sano,Y., Mazda,T. and Tomita,M.  
TITLE Purification and characterization of a novel hyaluronan-binding protein (PHBP) from human plasma: it has three EGF, a kringle and a serine protease domain, similar to hepatocyte growth factor activator  
J. Biochem. 119 (6), 1157-1165 (1996)  
JOURNAL 96425001  
MEDLINE 8827452  
PUBMED  
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gisseq 179630] from the original journal article.  
This sequence comes from Fig. 3B.  
FEATURES  
source  
1..2408  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1..2408  
/gene="hyaluronan-binding protein, PHBP"  
97..11779  
/gene="hyaluronan-binding protein, PHBP"  
/note="hepatocyte growth factor activator homolog; Method: conceptual translation with partial peptide sequencing; This sequence comes from Fig. 3B; PHBP"  
/codon\_start=1  
/product="hyaluronan-binding protein"  
/protein\_id="AAB46909.1"  
/db\_xref="GI:1836159"  
/translation="MFARMSDLHVLMLALVGKTAFCGFLSLSLESIDPDWTDPQVDY SYEDYNOEENTSSLTTHAENPDVYTEDADPCOPNCEHGGDLVHGSTFTCSCLAP PSGNCKQVONTCKDNPCGQCLITOSPYPYRCVKHPYTGPSOVPVCPNPO NGATCSHRKRSKFTACPDQFKFCEIGSDDCYVDGYSYRGKNRTVNOHACLYW NSHLLQENYMFDEATHGIGEHNFRCNPDADEKPCWCFIKVTDNKVWEYCDVSAQDVAPPEESPTEPKLPFGDSCKTEIAERKIKRIYGGFSKTAGKHWPQASLOSS LPLTISMPQGHFCGALTHPCWVLTAACTD IKTRHLKVLGQDLKKEEFHQSFV EKIIFYSHYNERDEIPNDIALKLKPDVGHCALESKYVKTCLPDGSPSPGSECHIS GWGVETGKGRLLDAKVLIAANTLCNSRLYDHMTDDSMICAGNLQKPGQDTCCQD SGGPLTCEKDGTYVYGVISWGLECGKRPVYTVQTKFLNWKATIKSESGF"

Query Match 100.0%; Score 1683; DB 9; Length 2408;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTTGCAGAGTCTGATCTCATCTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60  
Db 97 ATGTTTGCAGAGTCTGATCTCATCTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 156

QY 61 GCCTGTGGTTCCTCCCTGATGCTTTATTGGAAGCCCTGGACCCAGACTGGACCCCTGAC 120  
Db 157 GCCTGTGGTTCCTCCCTGATGCTTTATTGGAAGCCCTGGACCCAGACTGGACCCCTGAC 216

QY 121 CAGTATGATACAGCTACGAGATTATATCAGGAGAGACACAGTAGACACACTTACC 180  
Db 217 CAGTATGATACAGCTACGAGATTATATCAGGAGAGACACAGTAGACACTTACC 276

QY 181 CATGCTGAGAATCTCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCACCCCAAC 240  
Db 277 CATGCTGAGAATCTCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCACCCCAAC 336

QY 241 CCTGTGAACAGGTGGGACTGCTCGTCCATGGGACACCTTCACATGACAGTGCCTG 300  
Db 337 CCTGTGAACAGGTGGGACTGCTCGTCCATGGGACACCTTCACATGACAGTGCCTG 396

QY 301 GCTCTTTTCTCTGGGAATAAGTGTGAGAAAGTCAAAATAGTGCAGAGCAAAACCCCATGT 360  
Db 397 GCTCTTTTCTCTGGGAATAAGTGTGAGAAAGTCAAAATAGTGCAGAGCAAAACCCCATGT 456

QY 361 GCGCGGGGCAATGCTCATACCCAGAGTCTCTCTACTACCGTGTGTGTAAACAC 420  
Db 457 GCGCGGGGCAATGCTCATACCCAGAGTCTCTCTACTACCGTGTGTGTAAACAC 516

QY 421 CCTTACACAGGTGCCAGCTGCTCCCAAGTGTCTCTGTATCGAGGCCAAACCCCTGCCAG 480  
Db 517 CCTTACACAGTCCAGCTGCTCCCAAGTGTCTCTGTATCGAGGCCAAACCCCTGCCAG 576

QY 481 AATGGGCTACTGCTCCCGGATTAAGCGAGATCCAAAGTTCACCTGTGCTGTCCCGAC 540  
Db 577 AATGGGCTACTGCTCCCGGATTAAGCGAGATCCAAAGTTCACCTGTGCTGTCCCGAC 636

QY 541 CAGTTTCAAGGGGAAATCTGTGAAATAGTTCTGATGACTGCTATGTTGGCGATGGCTAC 600  
Db 637 CAGTTTCAAGGGGAAATCTGTGAAATAGTTCTGATGACTGCTATGTTGGCGATGGCTAC 696

QY 601 TCTTACCGAGGAAATGAATAGGACACTCAACAGCATGGTGCCTTTACTGGAACCTCC 660  
Db 697 TCTTACCGAGGAAATGAATAGGACACTCAACAGCATGGTGCCTTTACTGGAACCTCC 756

QY 661 CACCTCTCTTTCAGGAGAAATTTACAACATGTTTATGGAGGATGCTGAAACCCATGGGATT 720  
Db 757 CACCTCTCTTTCAGGAGAAATTTACAACATGTTTATGGAGGATGCTGAAACCCATGGGATT 816

QY 721 GGGGAACACAAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGTGTATTATAA 780  
Db 817 GGGGAACACAAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGTGTATTATAA 876

QY 781 GTTACCAATGACAAGGTGAATGGGAATACTGTGATGTCCTCAGCGTCTCAGCCAGAC 840  
Db 877 GTTACCAATGACAAGGTGAATGGGAATACTGTGATGTCCTCAGCGTCTCAGCCAGAC 936

QY 841 GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCCGGGTGTTGACTCC 900  
Db 937 GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCCGGGTGTTGACTCC 996

QY 901 TGTGGAAGACTGAGATAGCAGAGAGGAGATCAGAGAACTATGGAAGCTTTAAGAGC 960  
Db 997 TGTGGAAGACTGAGATAGCAGAGAGGAGATCAGAGAACTATGGAAGCTTTAAGAGC 1056

QY 961 ACGGGGGCAAGCACCCCATGGCAGCGCTCCCTCCAGTCTCTCGTGCCTCTGACCATCTCC 1020  
Db 1057 ACGGGGGCAAGCACCCCATGGCAGCGCTCCCTCCAGTCTCTCGTGCCTCTGACCATCTCC 1116

QY 1021 ATGCCCCAGGGCACTTCTGTGGTGGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080  
Db 1117 ATGCCCCAGGGCACTTCTGTGGTGGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1176

QY 1081 GCCCACTGCACGACATAAAAAACACAGACATCTAAAGGTGGTGTAGGGGACCCAGGACCTG 1140  
Db 1177 GCCCACTGCACGACATAAAAAACACAGACATCTAAAGGTGGTGTAGGGGACCCAGGACCTG 1236

QY 1141 AAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAGATATTCAAGTAGAGCCAC 1200  
Db 1237 AAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAGATATTCAAGTAGAGCCAC 1296

QY 1201 TACAATGAAAGAGATGAGATTCCCCACAAATGATATTGCAATGCTCAAGTTAAAGCCAGTG 1260  
Db 1297 TACAATGAAAGAGATGAGATTCCCCACAAATGATATTGCAATGCTCAAGTTAAAGCCAGTG 1356



QY 1261 GATGGTCACTGTGCTCTAGATCCAAATACGTGAGACGTGTGTGCTTGCCTGATGGGTCC 1320  
DB 1357 GATGGTCACTGTGCTCTAGATCCAAATACGTGAGACGTGTGTGCTTGCCTGATGGGTCC 1416  
QY 1321 TTTCCCTCTGGGAGTGAGTGCACATCTCTGGCTGGGGTGTACAGAAAACAGAAAAGGG 1380  
DB 1417 TTTCCCTCTGGGAGTGAGTGCACATCTCTGGCTGGGGTGTACAGAAAACAGAAAAGGG 1476  
QY 1381 TCCCGCAGCTCTGATGCCAAAGTCAAGCTGATTCGCAACACATTTGTGCAACTCCGCG 1440  
DB 1477 TCCCGCAGCTCTGATGCCAAAGTCAAGCTGATTCGCAACACATTTGTGCAACTCCGCG 1536  
QY 1441 CAACTCTATGACCAATGATGATGACAGTATGATCTGTGCGAGGAAATCTTCAGAAAACCT 1500  
DB 1537 CAACTCTATGACCAATGATGATGACAGTATGATCTGTGCGAGGAAATCTTCAGAAAACCT 1596  
QY 1501 GGGCAAGACACCTGCCAGGGTGACTCTGGAGGCCCTGACTGTGAGAGACGGCACC 1560  
DB 1597 GGGCAAGACACCTGCCAGGGTGACTCTGGAGGCCCTGACTGTGAGAGACGGCACC 1656  
QY 1561 TACTAGCTCTATGGGATAGTGGCTGGGCTGGAGTGGGAGAGGCGCAGGGGTCTAC 1620  
DB 1657 TACTAGCTCTATGGGATAGTGGCTGGGCTGGAGTGGGAGAGGCGCAGGGGTCTAC 1716  
QY 1621 ACCAAGTACCAAAATCTGAAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680  
DB 1717 ACCAAGTACCAAAATCTGAAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1776  
QY 1681 TAA 1683  
DB 1777 TAA 1779

RESULT 4  
AX409021 3008 bp DNA linear PAT 14-JUN-2002  
LOCUS Sequence 1668 from Patent WO0229103.  
DEFINITION AX409021  
ACCESSION AX409021  
VERSION AX409021.1 GI:21441726  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J. G.  
TITLE Gene expression profiles in liver cancer  
JOURNAL Patent: WO 0229103-A 1668 11-APR-2002;  
GENE LOGIC INC (US)  
FEATURES  
source 1..3008  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="EMBL/GenBank Accession No. D49742"

ORIGIN  
Query Match 100.0%; Score 1683; DB 6; Length 3008;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGTTTGGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAAGACA 60  
DB 97 ATGTTTGGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAAGACA 156  
QY 61 GCCTGTGGGTCTCCCTGATGCTTTTATGGAAAGCTTGACCCAGACTGGACCCCTGAC 120  
DB 157 GCCTGTGGGTCTCCCTGATGCTTTTATGGAAAGCTTGACCCAGACTGGACCCCTGAC 216  
QY 121 CAGTATGATTACAGCTACGAGGATTAATCAGAGAGAAACACCAAGTAGCACCTTACC 180  
DB 217 CAGTATGATTACAGCTACGAGGATTAATCAGAGAGAAACACCAAGTAGCACCTTACC 276

QY 181 CATGCTCAGAAATCCTGACTGTTACTACACTAGAGACCAAGCTGATCCATGCCAGCCCAAC 240  
DB 277 CATGCTCAGAAATCCTGACTGTTACTACACTAGAGACCAAGCTGATCCATGCCAGCCCAAC 336  
QY 241 CCCTGTGAACACAGGTGGGAGTGCCTCGTCCATGGGAGCACCTTCAATGACAGCTGCTG 300  
DB 337 CCCTGTGAACACAGGTGGGAGTGCCTCGTCCATGGGAGCACCTTCAATGACAGCTGCTG 396  
QY 301 GCTCCTTCTCTGGGANTAGTGTCAAGAAAGTCCAAATACGTGCGAGGACCAACCCATGT 360  
DB 397 GCTCCTTCTCTGGGANTAGTGTCAAGAAAGTCCAAATACGTGCGAGGACCAACCCATGT 456  
QY 361 GGGCGGGGCCAATGTCTCTCATACCCAGAGTCTCCCTACTACCCGCTGTGTCTTAAACAC 420  
DB 457 GGGCGGGGCCAATGTCTCTCATACCCAGAGTCTCCCTACTACCCGCTGTGTCTTAAACAC 516  
QY 421 CCCTTACACAGGTGCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGACCAACCCCTGCCAG 480  
DB 517 CCCTTACACAGGTGCCAGTCTCTCCCAAGTGGTTCCTGTATGAGGACCAACCCCTGCCAG 576  
QY 481 AATGGGCTACCTGCTCCCGGCATTAAGCGAGATCCAAAGTTCACCTGTGCTGTCTCCGAC 540  
DB 577 AATGGGCTACCTGCTCCCGGCATTAAGCGAGATCCAAAGTTCACCTGTGCTGTCTCCGAC 636  
QY 541 CAGTTCAAGGGGAAATTTCTGTGAATAGGTTCTGATGACTGCTATGTTGGCGATGCGTAC 600  
DB 637 CAGTTCAAGGGGAAATTTCTGTGAATAGGTTCTGATGACTGCTATGTTGGCGATGCGTAC 696  
QY 601 TCTTACCGAGGAAATGAATAGACAGTCAACAGCATCGTGCCTTTACTGCGAACTCC 660  
DB 697 TCTTACCGAGGAAATGAATAGACAGTCAACAGCATCGTGCCTTTACTGCGAACTCC 756  
QY 661 CACCTCTCTTGCAAGGAAATTAACAATTTTATGAGGATGCTGAAACCCATGGGATT 720  
DB 757 CACCTCTCTTGCAAGGAAATTAACAATTTTATGAGGATGCTGAAACCCATGGGATT 816  
QY 721 GGGGAAACACAAATTTCTGCAGAACCCAGATCGCGAGAAAGCCCTGCTGCTTTATTAAA 780  
DB 817 GGGGAAACACAAATTTCTGCAGAACCCAGATCGCGAGAAAGCCCTGCTGCTTTATTAAA 876  
QY 781 GTTACCAATGACAGGTGAAATGGGAATCTGTGATGCTCAGCCTGCTCAGCCCAAGGAC 840  
DB 877 GTTACCAATGACAGGTGAAATGGGAATCTGTGATGCTCAGCCTGCTCAGCCCAAGGAC 936  
QY 841 GTTCCCTACCCAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTGACTCC 900  
DB 937 GTTCCCTACCCAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTGACTCC 996  
QY 901 TGTGAAAGACTGAGATAGCAGAGAGGAAGATCAAGAGAAATCTATGGAGGCTTTAAGAGC 960  
DB 997 TGTGAAAGACTGAGATAGCAGAGAGGAAGATCAAGAGAAATCTATGGAGGCTTTAAGAGC 1056  
QY 961 ACGCGGGCGAAGCAACCATGGGAGGCTCCCTCAGTCTCGCTGCTGCTGACCACTCC 1020  
DB 1057 ACGCGGGCGAAGCAACCATGGGAGGCTCCCTCAGTCTCGCTGCTGCTGACCACTCC 1116  
QY 1021 ATGCCCCAGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080  
DB 1117 ATGCCCCAGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1176  
QY 1081 GCCCACTGCAACGACATAAAGAACAGACATCTTAAAGTGGTGTGCTAGGGGACCGAGGACTG 1140  
DB 1177 GCCCACTGCAACGACATAAAGAACAGACATCTTAAAGTGGTGTGCTAGGGGACCGAGGACTG 1236  
QY 1141 AAGAAAGAGAAATTTCTATGAGCAGAGCTTTAGGTTGGAGAGATATTCAAGTACAGCCAC 1200  
DB 1237 AAGAAAGAGAAATTTCTATGAGCAGAGCTTTAGGTTGGAGAGATATTCAAGTACAGCCAC 1296  
QY 1201 TACAATGAAGAGATGAGATTCCCAACATGATATTGCAATGCTCAAGTTGAAGCCAGTG 1260  
DB 1297 TACAATGAAGAGATGAGATTCCCAACATGATATTGCAATGCTCAAGTTGAAGCCAGTG 1356  
QY 1261 GATGGTCACTGTGCTCTAGAAATCCAAATACGTGAGAGATGTTGCTGCTGATGGGTCC 1320

1357	GATGTCAC	TGTGCTCTAGAA	TCCAAATCAGTGAAGACTGTGTGCTTGCTCTGATGGGTCC	1416	
1321	TTTCCCTCTGGAGTGAGTGCCACATCTCTGGCTGGGGTGTACAGAAACAGAAAGGG	1380			
1417	TTTCCCTCTGGAGTGAGTGCCACATCTCTGGCTGGGGTGTACAGAAACAGAAAGGG	1476			
1381	TCCGCCAGCTCCTGGATGCCAAAGTCAAGCTGATTGCCAACACATTTGTGCAACTCCCCG	1440			
1477	TCCGCCAGCTCCTGGATGCCAAAGTCAAGCTGATTGCCAACACATTTGTGCAACTCCCCG	1536			
1441	CAACTCTATGACACATGATTGATGACAGTATGATCTGTGCGAGGAAATTTTCAGAAAACCT	1500			
1537	CAACTCTATGACACATGATTGATGACAGTATGATCTGTGCGAGGAAATTTTCAGAAAACCT	1596			
1501	GGCGAAGACACTGCCAGAGTGACTCTGGAGGCCCTGACCTGTGAGAAGACGCGACCC	1560			
1597	GGCGAAGACACTGCCAGAGTGACTCTGGAGGCCCTGACCTGTGAGAAGACGCGACCC	1656			
1561	TACTACGCTATGGGATAGTAGCTGGGGCTCTGGAGTGTGGAAAGAGGCCAGGGGTCTAC	1620			
1657	TACTACGCTATGGGATAGTAGCTGGGGCTCTGGAGTGTGGAAAGAGGCCAGGGGTCTAC	1716			
1621	ACCCAAGTTACCAAAATTCCTGAATTGGATCAAAAGCCACCATCAAAAGTCAAAAGTGGCTTC	1680			
1717	ACCCAAGTTACCAAAATTCCTGAATTGGATCAAAAGCCACCATCAAAAGTCAAAAGTGGCTTC	1776			
1681	TAA	1683			
1777	TAA	1779			

RESULT 5	
HUMHGFGAL	
LOCUS	3008 bp mRNA linear PRI 10-FEB-1999
DEFINITION	Human mRNA for HGF activator like protein, complete cds.
ACCESSION	D49742
VERSION	D49742.1 GI:736706
KEYWORDS	HGF activator like protein; serin protease.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE	1 (bases 1 to 3008)
AUTHORS	Kitamura,N.
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 3008)
AUTHORS	Kitamura,N.
TITLE	Direct Submission
JOURNAL	Submitted (17-MAR-1995) Naomi Kitamura, Institute for Liver Research, Kansei Medical University; Moriguchi, Osaka 570, Japan (Tel.:06-992-1001(ex.2530). Fax:06-994-6099)

FEATURES	source
Location/Qualifiers	
1..3008	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
97..1779	
/codon_start=1	
/product="HGF activator like protein"	
/protein_id="BAA08576.1"	
/db_xref="GI:134539"	
/translation="MEARMSDLHLVLLMALVGTACGFSLSMLLESLSDDPDWTDQDQYD SYEDYQNEBNTSLTHAENSDNHWYTEDAQCPQPCRGHGDCLVHGSTFTFCSCILAP FSNGKQKQVNTCKDMPGCRGQLITQSPFYCYGVCKHPYSCSVVPCVPCNPQC NGATCRKHRSKFTCAQCPQFGKTCETGSDCYGVSGYSGKMRVYNQAHCLYQ NSHLLQENNYNMEDEATHTGIEHNFNCEADDEKPCWFIKYNDKWKVEYCDVAS SAQDVAYPESPETPKLPGFDSCKTELAEKIKLYGGFKSTAGKHPQWASLOSS LEPTISMPQHFCCGLIHPCWVITAACTDIKRLHKVVLGQDILAKEEFHPQSRV EKWFTYSHYNERDEIPIHNDIALKLFPVDGHGCALESKYVATVCLPGSGTFSGSECHIS GKATVSGKSRQLLQALKEIANTLCSNQLYDHMDIDDSMIAGNLLKPGQDTCCGD SGGLFCEKDGTGYVYGVISWGLECCRGCPGVYQVTKFLNWKATIKSESGF"	
2312	
polvA site	

[illegible]



/note="KR; Region: Kringle domain"  
/db\_xref="CDD:smart00130"  
1017. 1730  
/note="Tryp\_SPC; Region: Trypsin-like serine protease"  
/db\_xref="CDD:smart00020"

ORIGIN

Query Match 99.9%; Score 1681.4; DB 9; Length 2251;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1682; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGTTTCCAGGAGTCTGATCTCCATGTTCTGCTTAATGGCTCTGGTGGGAAAGACA 60  
Db 81 ATGTTTCCAGGAGTCTGATCTCCATGTTCTGCTTAATGGCTCTGGTGGGAAAGACA 140  
Qy 61 GCCTGGGTTCTCCCTGATGCTTTTATGGAAGCCTGGAACCAAGTGGACCCCTGAC 120  
Db 141 GCCTGGGTTCTCCCTGATGCTTTTATGGAAGCCTGGAACCAAGTGGACCCCTGAC 200  
Qy 121 CAGTATGATACAGTACAGGAGTATTAATCAGGAAGAGAACCAAGTACCACTTACC 180  
Db 201 CAGTATGATACAGTACAGGAGTATTAATCAGGAAGAGAACCAAGTACCACTTACC 260  
Qy 181 CATGCTGAGAACTCTGACTGCTATACACTGAGGACCAAGTATGATCCAGCCCAAC 240  
Db 261 CACGCTGAGAACTCTGACTGCTATACACTGAGGACCAAGTATGATCCAGCCCAAC 320  
Qy 241 CCCTGTGAACACGCTGGGAGTCTGCTGCTCATGGGAGACCTTCAATCAGTCCGCTG 300  
Db 321 CCCTGTGAACACGCTGGGAGTCTGCTGCTCATGGGAGACCTTCAATCAGTCCGCTG 380  
Qy 301 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTCGAAAGCAACCCATGT 360  
Db 381 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTCGAAAGCAACCCATGT 440  
Qy 361 GGCCGGGGCAATCTCATACCCAGAGTCTCCCTACTACCGTCTGCTGTGAACAC 420  
Db 441 GGCCGGGGCAATCTCATACCCAGAGTCTCCCTACTACCGTCTGCTGTGAACAC 500  
Qy 421 CTTTACACAGTCCAGTCTCCCAAGTGTCTCTGATGACGCTATGTCAGGCAAAACCCCTGCCAG 480  
Db 501 CTTTACACAGTCCAGTCTCCCAAGTGTCTCTGATGACGCTATGTCAGGCAAAACCCCTGCCAG 560  
Qy 481 AATGGGCTACTCTCTCCGCAATAGGGAGATCCAGTTCACCTGCTGCTGCTCCGAC 540  
Db 561 AATGGGCTACTCTCTCCGCAATAGGGAGATCCAGTTCACCTGCTGCTGCTCCGAC 620  
Qy 541 CAGTTCAAGGGGAAATCTGTGAATAGTGTCTGATGACGCTATGTCAGTATGTCGATGGCTAC 600  
Db 621 CAGTTCAAGGGGAAATCTGTGAATAGTGTCTGATGACGCTATGTCAGTATGTCGATGGCTAC 680  
Qy 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCAGTCCGCTCTTACTGGAATCC 660  
Db 681 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCAGTCCGCTCTTACTGGAATCC 740  
Qy 661 CACCTCTCTTTCAGGAGAAATCAACATGTTTATGGAGGATGCTGAAACCCATGGGATT 720  
Db 741 CACCTCTCTTTCAGGAGAAATCAACATGTTTATGGAGGATGCTGAAACCCATGGGATT 800  
Qy 721 GGGGAACACAAATTTCTGAGAAACCCAGATCGGAGCAAAAGCCCTGCTTTATTTAAA 780  
Db 801 GGGGAACACAAATTTCTGAGAAACCCAGATCGGAGCAAAAGCCCTGCTTTATTTAAA 860  
Qy 781 GTTACCAATGACAAAGTGAATGGAATACGTGATGCTCAGGCTGCTCAGCCCAAGAC 840  
Db 861 GTTACCAATGACAAAGTGAATGGAATACGTGATGCTCAGGCTGCTCAGCCCAAGAC 920  
Qy 841 GTTGCTTACCCAGAGAAAGCCCACTGAGCAATCAACCAAGCTTCGGGGTTGATCC 900  
Db 921 GTTGCTTACCCAGAGAAAGCCCACTGAGCAATCAACCAAGCTTCGGGGTTGATCC 980  
Qy 901 TGTGAAAGACTGAGATAGCAGAGGAGATCAAGAAATCTATGAGGCTTTAAGACG 960

Db 981 TGTGAAAGACTGAGATAGCAGAGAGGAAGATCAAGAGAAATCTATGGAGGCTTTAAGAGC 1040  
Qy 961 ACGGGGGCAAGCACCACCATGSCAGCGTCCCTCAGTCTCTGCTGCTCTGACCATCTCC 1020  
Db 1041 ACGGGGGCAAGCACCACCATGSCAGCGTCCCTCAGTCTCTGCTGCTCTGACCATCTCC 1100  
Qy 1021 ATGCCCCAGGCGCACTTCTGTGTGGGCGCTGATCCACCCCTGCTGGTGTCTCACTGCT 1080  
Db 1101 ATGCCCCAGGCGCACTTCTGTGTGGGCGCTGATCCACCCCTGCTGGTGTCTCACTGCT 1160  
Qy 1081 GCCCACTGCACCGACATAAAAAACCCAGACATCTAAAGTGTGTGTAGGGAGCCAGGACCTG 1140  
Db 1161 GCCCACTGCACCGACATAAAAAACCCAGACATCTAAAGTGTGTGTAGGGAGCCAGGACCTG 1220  
Qy 1141 AAGAAGAGAAATTTCTATGACGACAGCTTTTAGGCTGGAGAGATATCAAGTACAGCCAC 1200  
Db 1221 AAGAAGAGAAATTTCTATGACGACAGCTTTTAGGCTGGAGAGATATCAAGTACAGCCAC 1280  
Qy 1201 TACATGAAAGAGATGAGATTCCCAATGATATTGCTCAAGTCTCAAGTCTCAAGTCTCAAGTCT 1260  
Db 1281 TACATGAAAGAGATGAGATTCCCAATGATATTGCTCAAGTCTCAAGTCTCAAGTCTCAAGTCT 1340  
Qy 1261 GATGTCTACTGTCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGGTCC 1320  
Db 1341 GATGTCTACTGTCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGGTCC 1400  
Qy 1321 TTTCCCTCTGGGAGTGAAGTGCACATCTCTGCTGGGCTGTACAGAAACAGGAAAGGG 1380  
Db 1401 TTTCCCTCTGGGAGTGAAGTGCACATCTCTGCTGGGCTGTACAGAAACAGGAAAGGG 1460  
Qy 1381 TCCGCCAGCTCTCGGATGCAAGTCAAGTGAATGCAACACTTTGTGCAACTCCGCG 1440  
Db 1461 TCCGCCAGCTCTCGGATGCAAGTCAAGTGAATGCAACACTTTGTGCAACTCCGCG 1520  
Qy 1441 CAACTCTATGACCAATGATGATGACAGTATGATGTCAGGAAATCTTTCAGAAACCT 1500  
Db 1521 CAACTCTATGACCAATGATGATGACAGTATGATGTCAGGAAATCTTTCAGAAACCT 1580  
Qy 1501 GGGGAAGACACTGCGAGGCTGACTCTGAGAGCCCTGACCTGTGAGAGAGAGCCAGGCTCTAC 1560  
Db 1581 GGGGAAGACACTGCGAGGCTGACTCTGAGAGCCCTGACCTGTGAGAGAGAGCCAGGCTCTAC 1640  
Qy 1561 TACTACGCTATGGGATAGTGAAGTGGGCTGAGTGTGGAGAGAGCCAGGCTCTAC 1620  
Db 1641 TACTACGCTATGGGATAGTGAAGTGGGCTGAGTGTGGAGAGAGCCAGGCTCTAC 1700  
Qy 1621 ACCCAAGTTACAAATCTCTGAATGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC 1680  
Db 1701 ACCCAAGTTACAAATCTCTGAATGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC 1760  
Qy 1681 TAA 1683  
Db 1761 TAA 1763

RESULT 7  
AX383955  
LOCUS  
DEFINITION  
Sequence 2 from Patent EP1182258.  
AX383955  
VERSION  
AX383955.1 GI:19577526  
KEYWORDS  
Homo sapiens (human)  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE  
AUTHORS  
Roemisch, J., Stoeck, H.A., Feussner, A., Lang, W., Weimer, T.,  
Becker, M., Nerlich, C. and Muth-Naumann, G.  
TITLE  
Mutants of the factor vii activating protease and methods for their  
detection  
JOURNAL  
Patent: EP 1182258-A 2 27-FEB-2002;  
Aventis Behring GmbH (DE)

FEATURES	Location/Qualifiers
source	1..1683
	.organism="Homo sapiens"
	/mol_type="unassigned DNA"
	/db_xref="taxon:9606"
ORIGIN	
Query Match	99.8%; Score 1679.8; DB 6; Length 1683;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 1681; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
QY	1 ATGTTTGCCAGGATGCTGTGATCTCCATGTTCTGCTGTTAATGCCTCTGCTGGGAAGACA 60
DB	1 ATGTTTGCCAGATGTCGATCTCCATGTTCTGCTGTTAATGGCTCTGTTGGGAAGACA 60
QY	61 GCCTGTGGGTTCTCCCTGATGTCCTTTATTGGAAGCCTCGAACCCAGACTGGACCCCTGCAC 120
DB	61 GCCTGTGGGTTCTCCCTGATGTCCTTTATTGGAAGCCTCGAACCCAGACTGGACCCCTGCAC 120
QY	121 CAGTATGATTACACTACGAGGAGTTAATAATCAGGAAGAGNAACCAAGTAGCACACTTACC 180
DB	121 CAGTATGATTACACTACGAGGAGTTAATAATCAGGAAGAGAACACCAAGTAGCACACTTACC 180
QY	181 CATGCTGAGAAATCCCTGATGCTGTACTACACTGAGGACCAAGCTGATCCATGCCAGGCCAAC 240
DB	181 CATGCTGAGAAATCCCTGATGCTGTACTACACTGAGGACCAAGCTGATCCATGCCAGGCCAAC 240
QY	241 CCCTGTGAACACGGTGGGACTGCTCGTCCATGGGAGCACCTTCACATGCAGCTGCCTG 300
DB	241 CCCTGTGAACACGGTGGGACTGCTCGTCCATGGGAGCACCTTCACATGCAGCTGCCTG 300
QY	301 GCTCCTTTCTCTGGGANAATAGCTGTCAGAAAGTGCAGAAATACGTGCAAGGACAAACCCATGT 360
DB	301 GCTCCTTTCTCTGGGANAATAGCTGTCAGAAAGTGCAGAAATACGTGCAAGGACAAACCCATGT 360
QY	361 GCGCGGGGCCAATGTCTCATATPACAGAGTCTCCCTACTACCGCTGTGCTGTGTAACAC 420
DB	361 GCGCGGGGCCAATGTCTCATATPACAGAGTCTCCCTACTACCGCTGTGCTGTGTAACAC 420
QY	421 CCTTTACAGAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGCAAGGCCAAACCCCTGCCAG 480
DB	421 CCTTTACAGAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGCAAGGCCAAACCCCTGCCAG 480
QY	481 AATGGGGGTACTGCTCTCCGGCATAAGCGGAGATCCAAGTTACCTGTGCTCTGCCGAC 540
DB	481 AATGGGGGTACTGCTCTCCGGCATAAGCGGAGATCCAAGTTACCTGTGCTCTGCCGAC 540
QY	541 CAGTTCAAGGGGAAATTTCTGTAATAGGTTCTTGATGACTGCTATGTTGGCGATGGCTAC 600
DB	541 CAGTTCAAGGGGAAATTTCTGTAATAGGTTCTTGATGACTGCTATGTTGGCGATGGCTAC 600
QY	601 TCTTACCAGGGGAAATGAATAGGACAGTCAAACAGCATGCGTGCCTTTACTTGGAACTCC 660
DB	601 TCTTACCAGGGGAAATGAATAGGACAGTCAAACAGCATGCGTGCCTTTACTTGGAACTCC 660
QY	661 CACTCTCTTTGACGAGAAATTAACAATGTTTATGGAGGATGCTGAAACCCATGGGATT 720
DB	661 CACTCTCTTTGACGAGAAATTAACAATGTTTATGGAGGATGCTGAAACCCATGGGATT 720
QY	721 GGGGAACACAATTTCTGCAGAAAACCCAGATGCGGACGAAAGCCCTGGTCTTTTATATAA 780
DB	721 GGGGAACACAATTTCTGCAGAAAACCCAGATGCGGACGAAAGCCCTGGTCTTTTATATAA 780
QY	781 GTTACCAATGACAAGGTGAAATGGGAATACTGTGATGCTCTCAGCCCTGCTCAGCCCGAGAC 840
DB	781 GTTACCAATGACAAGGTGAAATGGGAATACTGTGATGCTCTCAGCCCTGCTCAGCCCGAGAC 840
QY	841 GTTGTCTACCCAGAGGAAGCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC 900
DB	841 GTTGTCTACCCAGAGGAAGCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC 900
QY	901 TGTGGAAGACTTGAGATAGCAGAGGAAGATCAAGAGAACTCTATGGAGGCTTTTAAGAGC 960

901	TGTGGAAGACTGAGATAGCAGAGAGGAAGATCAAGAGAACTATGAGAGCGTTTAAAGAC	960
961	ACGGCGGGCAAGCACCACCATGCGAGGGCGTCCCTCCAGTCTCGCTGCTCTGACCATCTCC	1020
961	ACGGCGGGCAAGCACCACCATGCGAGGGCGTCCCTCCAGTCTCGCTGCTCTGACCATCTCC	1020
1021	ATGCCCCAGGGCCACATCTCTGTGTGTGGGCGCTGATCCACCCCTGCTGGTCTCAGCTGCT	1080
1021	ATGCCCCAGGGCCACATCTCTGTGTGTGGGCGCTGATCCACCCCTGCTGGTCTCAGCTGCT	1080
1081	GCCCACTGCACCCAGACATAAAAAACAGACATCTAAAGGTGGTGTAGGGGAGACAGACCTG	1140
1081	GCCCACTGCACCCAGACATAAAAAACAGACATCTAAAGGTGGTGTAGGGGAGACAGACCTG	1140
1141	AAGAAAGAGAAATTTATGAGCAGAGCTTTAGGTGTGGAGAGATATTCAGTACAGCCAC	1200
1141	AAGAAAGAGAAATTTATGAGCAGAGCTTTAGGTGTGGAGAGATATTCAGTACAGCCAC	1200
1201	TACAATGAAGAGATGAGATTCCCAACAATGATATTGCAATGCTCAAGTTAAAGCCAGTG	1260
1201	TACAATGAAGAGATGAGATTCCCAACAATGATATTGCAATGCTCAAGTTAAAGCCAGTG	1260
1261	GATGGTCACTGTGCTTGTAGAAATCCAAATACGTGAAGACGTGTGTGCTGTGATGGGTC	1320
1261	GATGGTCACTGTGCTTGTAGAAATCCAAATACGTGAAGACGTGTGTGCTGTGATGGGTC	1320
1321	TTTCCCTCTGGGAGTGAGTGCCACATCTCTGCTGGGTGTTACAGAAACAGGAAAGGG	1380
1321	TTTCCCTCTGGGAGTGAGTGCCACATCTCTGCTGGGTGTTACAGAAACAGGAAAGGG	1380
1381	TCCGCCAGCTCTCTGGATGCCAAAGTCAAGCTGATTGGCCAACTTTGTGCAACTCCCGC	1440
1381	TCCGCCAGCTCTCTGGATGCCAAAGTCAAGCTGATTGGCCAACTTTGTGCAACTCCCGC	1440
1441	CAACTCTATGACCACATGATTGATGACAGTATGATCTGTGACGAGGAATCTTCAGAAACCT	1500
1441	CAACTCTATGACCACATGATTGATGACAGTATGATCTGTGACGAGGAATCTTCAGAAACCT	1500
1501	GGGCAAGACACCTGCCAGGCTGACTCTGGAGGCCCTCCCTGACCTGTGAGAAGGACGGCAC	1560
1501	GGGCAAGACACCTGCCAGGCTGACTCTGGAGGCCCTCCCTGACCTGTGAGAAGGACGGCAC	1560
1561	TACTACGTCTATGGATAGTAGACTGGGCGCTGGAGTGTGGGAAGAGCCAGGGGTCTAC	1620
1561	TACTACGTCTATGGATAGTAGACTGGGCGCTGGAGTGTGGGAAGAGCCAGGGGTCTAC	1620
1621	ACCCAAGTTACCAAAATTCCTGAATTCGATCAAAAGCCACCATCAAAGTGAAGTGGCTTC	1680
1621	ACCCAAGTTACCAAAATTCCTGAATTCGATCAAAAGCCACCATCAAAGTGAAGTGGCTTC	1680
1681	TAA 1683	
1681	TAA 1683	
RESULT 8	BD177603	
LOCUS	BD177603	
DEFINITION	Detection method with the use of factor VII activating protease mutant and specific antibody.	
ACCESSION	BD177603	
VERSION	BD177603.1 GI:30014865	
KEYWORDS	JP 2002291486-A/2.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	Roemisch, J., Stoeck, H.A., Feussner, A., Lang, W., Weimer, T., Becker, M., Nerlich, C. and Naumann, G.M.	
TITLE	Detection method with the use of factor VII activating protease mutant and specific antibody	
JOURNAL	Patent: JP 2002291486-A 2 08-OCT-2002;	









[illegible]

RESULT 10	BC031775	2119 bp	linear	ROD 04-OCT-2003
BC031775	Mus musculus hyaluronin acid binding protein 2, mRNA (cDNA clone			
LOCUS	MG:28705 IMAGE:4242577), complete cds.			
DEFINITION				

RC031775.1 GI:21594554  
MUS musculus (house mouse)  
MUS musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 2119)  
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.

Klaesner, R.D., Collins, F.S., Wegner, L., Shenneman, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,

McKernan K. J., Malek, J. A., Gunaratne, P. H., Richards, S., Wexley, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hulyk, S. W., Villalon, D. K., Hale, S., Garcia, A. M., Gay, L. J., Hulyk, S. W., Villalon, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A. C., Shevchenko, Y., Bouffard, G. G., Blakesley, R. W., Touchman, J. W., Green, E. D., Dickinson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M., Butterfield, Y. S., Krzywiński, M. I., Skalski, U., Smalley, D. E.,

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
MEDLINE 22388357  
FUEMED 12477932  
REFERENCE 2 (bases 1 to 2119)

AUTHORS	Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (08-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 1IA03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>

COMMENT
<p>Contact: MGC help desk  Email: cgapbs@mail.nih.gov  Tissue Procurement: Jeffrey E. Green, M.D.  CDNA Library Preparation: Life Technologies, Inc.  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)</p>

QY 1541 CCTGTGAGAGGAGCGGACCTACTAGCTTATGGGATAGTGGGCGCTGGAGTGTG 1600  
 Db 1470 CCTGTGAGAGGATGGAAGTCTTACTAGCTTACGGGATTTAAGCTGGGCGCAGGATGTG 1529  
 QY 1601 GGAAGAGGCGAGGGTCTACACCCAGTTACCAATTCCTGAATTGGATCAAGCCACCA 1660  
 Db 1530 GGAAGAGGCGAGGCTTACACTCAAGTCCCAAGTCTCTGAATTGGAATAAGACCA 1589  
 QY 1661 TCAAAAGTGAAGTGGCTTCTAA 1683  
 Db 1590 TGCACAGGAGGCTGGCCTCTGA 1612

RESULT 11  
 AL390197 188349 bp DNA linear PRI 29-AUG-2001  
 LOCUS Human DNA sequence from clone RP11-190F19 on chromosome 10,  
 DEFINITION complete sequence.  
 ACCESSION AL390197  
 VERSION AL390197.19 GI:15387800  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Tracey, A.  
 Direct Submission  
 Submitted (29-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk  
 On Aug 31, 2001 this sequence version replaced gi:14596344.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em:, EMBL; Sw:,  
 SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 10, constructed by the Sanger Centre Chromosome 10  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-190F19 is from the library RPI-11.1 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTER: pBACE3.6  
 This sequence is the entire insert of clone RP11-190F19 The true  
 right end of clone RP11-481H2 is at 111825 in this sequence.

FEATURES  
 source  
 1. 188349  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-190F19"  
 /clone\_lib="RPI-11.1"  
 22142..22187  
 /notes="Sequence from overlapping clone RP11-481H2  
 (AL133482). Assembly confirmed by restriction digest."

misc\_feature  
 ORIGIN

QY 461 GCAGCCAAACCCCTGCCAGATGGGCTACCTGCCCGCATAGCGGAGATCCAAAGT 520  
 Db 390 GCAGCCAAACCCCTGCCAGATGGGCTACCTGCCCGCATAGCGGAGATCCAAAGT 449  
 QY 521 TCACCTGTGCTGTCCCGCAGTTCAAGGGGAATTCGTGAATAGTCTCATGACT 580  
 Db 450 TTACTGTGCTGTCCCGCAGTTCAAGGGGAATTCGTGAATAGTCTCATGACT 509  
 QY 581 GCTATGTTGGGATGCTACTCTTACCGAGGAAATCAATAGGACAGTCAACAGCATG 640  
 Db 510 GTTATGTCGGTATGGCTACTCTTACCGAGGCAAGTGAAGACAGTCAACAGAAC 569  
 QY 641 CGTGCCTTACTGGAACCTCCACCTCTCTGCGAGGAATTAACAATGTTATGAGG 700  
 Db 570 CATGCTTTACTGGAACCTCCACCTCTCTGCGAGGACTTAAACATGTTTATGAGG 629  
 QY 701 ATGCTGAAACCCATGGGATTGGGGAACACAATTTCTGCAGAAACCCAGATCGGACGAA 760  
 Db 630 ATGCAGAGACCCACGGATCGCAGAGCACAACTTCTGCAGAAACCCAGATCGGACCA 689  
 QY 761 AGCCCTGTGCTTTTAAAGTATCAATGACAAGTGAATGGGAATPACTGTGATGCT 820  
 Db 690 AACCTGTGTTTCTGCAAGTGAACAGTGAAGGTGAATGGGAATPACTGTGATGCT 749  
 QY 821 CAGCCTGTCTCAGCCAGGACGTTGCTCTACCCAGAGGAAAGCCCACTGAGCCATCAACA 880  
 Db 750 CAGTCTGTCCAGTCTGACACCCCTTACCCAGTGGAAAGCTTCTGGAGCTGTGATGG 809  
 QY 881 AGCTTCCGGGTTTGAATCTCTGTGAAGAAGATGAGATAGCAGAGGAAGATCAAGAGAA 940  
 Db 810 AGCTCCAGGGTTTGAATCTCTGTGAAGAAGATGAGATAGCAGAGGAAGATCAAGAGAA 869  
 QY 941 TCTATGAGGCTTTAAGAGCAGCGGCGGCAAGCACCCTGAGCGCTCCCTCCAGTCT 1000  
 Db 870 TCTAGGGGCTTTAAGAGCAGCGGCGGCAAGCACCCTGAGCGCTCCCTCCAGTCT 929  
 QY 1001 CGCTGCTCTGACCTTCCATCTCCCGGCGGCTCTCTGTGGTGGGCGCTGATCCACC 1060  
 Db 930 CACTGCGGTGACCACTCCATCTCCCGGCGGCTCTCTGTGGTGGGCGCTGATCCACC 989  
 QY 1061 CTTGTGGTGTCTACTGTCGCCCTCAGCGATCAATCAAAACAGATCTTAAGGTGG 1120  
 Db 990 CTTGTGGTGTCTACTGTCGCCCTCAGCGATCAATCAAAACAGATCTTAAGGTGG 1049  
 QY 1121 TGCTAGGGGACCGAGCACTTGAAGAAGAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGA 1180  
 Db 1050 TACTAGGGGATCAGGACCTTGAAGAAGAGAGACAGATCCCATGAACAGACCTTCAGGGTGGAAA 1109  
 QY 1181 AGATATTCACTACAGCCACTACATGAAGATGAGATCCCAAGATGATTTGCAT 1240  
 Db 1110 AAATACTGAAGTACAGTCAAGTATGAAGAAGATGAGATCCCAAGATGATTTGCAT 1169  
 QY 1241 TGCTCAAGTTAAAGCCAGTGGATGGTCACTGTGCTCTAGAAATCCAAATACGTAAGACTG 1300  
 Db 1170 TGCTCAAGTTAAAGCCAGTGGTGGTCACTGTGCTCTAGAAATCCCAAGATGATTTGCAT 1229  
 QY 1301 TGCTGTGCTGATGGTCTTCCCTCTGGGAGTGAAGTGCACATCTCTGGCTGGGCTG 1360  
 Db 1230 TATGTTTGCCAGCGACCCCTTCCCTCTGGAACTGAGTGCACATCTCTGGCTGGGCTG 1289  
 QY 1361 TTACAGAAACAGAAAGGGTCCCGCAGCTCTCTGATGCCAAAGTCAAGCTGATTTGCCA 1420  
 Db 1290 TTACAGAAACAGAAAGGGTCCCGCAGCTCTCTGATGCCAAAGTCAAGCTGATTTGCCA 1349  
 QY 1421 ACATTTGTGCAATCTCCCGCAGTCTTATGACACATGATGATGACAGTATGATCTGTG 1480  
 Db 1350 ACCCTTTGTGCAATCTCCCGCAGTCTTATGACACATGATGATGACAGTATGATCTGTG 1409  
 QY 1481 CAGGAATCTTCAGAACTCTGGCAGAGACCTGCGAGGGTCACTCTGGAGGCCCTCTGA 1540  
 Db 1410 CGGGGAACCTTCAGAAAGCCCGGATCAGACCTTCCAGGGTCACTCTGGAGGCCCTCTGA 1469





misc\_feature 136063..181391  
/note="assembly\_fragment"  
181492..204214  
/note="assembly\_fragment"  
clone end:T7  
vector\_side:right"

ORIGIN

Query Match 9.9%; Score 167.4; DB 2; Length 204214;  
Best Local Similarity 78.2%; Pred. No. 2.7e-38;  
Matches 201; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 838 GACGTTGCTACCCAGAGGAAGCCCACTAGACCATCAACCAAGCTTCGGGGTTTCAC 897  
Ddb 142187 GACACCCCTAACCCAGTGGAAAGCCTTCTGGAGCCTGTGATGAGCTGCACGGGTTCAG 142246

QY 898 TCCTGTGGAAAGACTGAGATAGCAGAGGAAGATCAAGAGATCTATGAGGCTTTAAG 957  
Ddb 142247 TCCGTGGGGAAGAAGGAGGTAGCTGAACACGCAAGTCAAGCGTATCTACGGGGCTTTAAG 142306

QY 958 AGCAGCGGGGCAAGCACCATCGCAGCGCTCCCTCCAGTCTCGCTCGCTCGACCATC 1017  
Ddb 142307 AGCAGCAGCAAGCACCCTGGCAGGTGTCCTCGACACCTCACTGCGGTGACCAACC 142366

QY 1018 TCCATGCCCCAGGCGCACTTCTGTGGTGGGGCGCTCATCCACCTGCTGGGTGCTCACT 1077  
Ddb 142367 TCCATGCCCCAAGGCGCACTTCTGTGGGCGGCGCCTGATCACCCCTGCTGGGTGCTCACT 142426

QY 1078 GCTGCCCACTGCACCGA 1094  
Ddb 142427 GCAGCCCACTGTACCGA 142443

RESULT 15  
AC106236/c

LOCUS  
DEFINITION  
AC106236.5 GI:30579797  
HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
1 (bases 1 to 224208)  
Muzny D, Marie E, Metzker M, Lee A, Abramson S, Adams C, Alder J, Allen C, Allen H, Albrechts S, Amin A, Anguiano D, Anyalebechi V, Ayagi A, Ayodeji M, Baca E, Baden H, Baldwin D, Bandaranaike D, Barber M, Barnstead M, Berahmed F, Biswal K, Blair J, Blankenburg K, Blyth P, Brown M, Bryant N, Buhay K, Burch P, Burrell K, Caldeiron E, Cardenas V, Carter K, Cavazos I, Ceasar H, Center A, Chacko J, Chavez D, Chen G, Chen R, Chen Y, Chen Z, Chu J, Cleveland C, Cockrell R, Cox C, Coyly M, Cree A, D'Souza L, Davila M, L, Davis C, Davy-Cartoll L, De Anda C, Dederich D, Delgado O, Denson S, Deramo C, Ding Y, Dinh H, Divya K, Draper H, Dugan-Rocha S, Dunn A, Durbin K, Duval B, Eaves K, Egan A, Escotto M, Eugene C, Evans C, Fallis I, Fan G, Fernandez S, Finley M, Flagg N, Forbes L, Foster M, Foster P, Fraser C, M, Gabisi A, Ganta R, Garcia A, Garner T, Garza M, Gebregorgis B, Geer K, Gill R, Grady M, Guerra W, Guevara W, Gunaratne P, Haaland W, Hamill C, Hamilton C, Hamilton K, Harvey Y, Havlak P, Hawes A, Henderson N, Hernandez J, Hernandez R, Hines S, Hladun S, L, Hodgson A, Hoques M, Hollins B, Howells S, Hulyk S, Hume J, Idlebird D, Jackson A, Jackson L, Jacob L, Jiang H, Johnson B, Johnson R, Jolivet A, Karpathy S, Kelly S, Kelly S, Khan Z, King L, Kovar C, Liu J, C, Kraft C, Kratt C, Lebow H, Levan J, Lewis L, Li Z, Liu J, Liu J, Liu W, Liu Y, London P, Longacre S, Lopez J, Lorushtewa L, Louised H, Lozada R, J, Lu X, Ma J, Maheshwari M, Mahindratne M, Mahmoud M, Malloy K, Mangum A, Mani

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokenkeme, O., Okwuonu, G., Olarunpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, X., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wleciyzk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 224208)  
Worley, K.C.  
Direct Submission  
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 224208)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 13, 2003 this sequence version replaced gi:24942489.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GKCO  
Center clone name: CH230-131J10  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 21390 bases at least Q40  
Consensus quality: 215800 bases at least Q30  
Consensus quality: 216989 bases at least Q20  
Estimated insert size: 222031; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 196130: contig of 196130 bp in length
* 196131 196230: gap of unknown length
* 196231 224208: contig of 27978 bp in length.
FEATURES
    source
        1..224208
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-131J10"
            2668..3129
                /note="clone boundary"
            clone_end:Sp6
            site:EcoRI
            end_sequence:BH346758"
            200382..201926
                /note="wgs_contig"
    misc_feature
    misc_feature
ORIGIN
Query Match          9.6%; Score 161; DB 2; Length 224208;
Best Local Similarity 76.7%; Pred. No. 2.4e-36;
Matches 197; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 838 GACGTTGCTACCCAGAGGAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTGAC 897
Db 71048 GATGCTGTATACCCATTGGGAGCCTTCAGGAACCTGTGATGGAGCTGCCAGGATTCGAC 70989
QY 898 TCCTGTGAAAGACTGAGATAGCAGAGAGGAAGATCAAGAGAATCTATGGAGGCTTTAAG 957
Db 70988 TCCTGCGGGAAGACAGAGATGACTGACTGACACACGCGGTCAAGCGCATCTACGGGGGCTTTAAG 70929
QY 958 AGCAGCGGGGCAAGCACCACCTGGCAGCGCTCCCTCCAGTCTCGCTGCTGACCATC 1017
Db 70928 AGCAGCGGGGCAAGCACCACCTGGCAGGTGTCCTTCGAGACCTCATTCGCGCTGACCACC 70869
QY 1018 TCCATGCCCCAGGGCCACTTCTGTGGTGGGGGGTGATCCACCCCTGCTGGGTGCTCACT 1077
Db 70868 TCCATGCCCCAAGGCCACTTCTGTGGGGGTTCCCTGATTCAACCCTTGCTGGGTGCTCACT 70809
QY 1078 GCTGCCCACTGCACCGA 1094
Db 70808 GCAGCCCACTGTACCGA 70792

Search completed: May 25, 2004, 09:00:30
Job time : 4486.5 secs

```

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 04:19:43 ; Search time 471 Seconds  
(without alignments)  
15179.867 Million cell updates/sec

Title: US-09-912-559-1  
Perfect score: 1683  
Sequence: 1 atgtttgcaggatgtctga.....aaagtgaagtggtttctaa 1683

Scoring table: IDENTITY\_NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_29Jan04:\*

- 1: Geneseqm1980s:\*
- 2: Geneseqm1980s:\*
- 3: Geneseqm2000s:\*
- 4: Geneseqm2001as:\*
- 5: Geneseqm2001bs:\*
- 6: Geneseqm2002s:\*
- 7: Geneseqm2003as:\*
- 8: Geneseqm2003bs:\*
- 9: Geneseqm2003cs:\*
- 10: Geneseqm2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1683	100.0	1683	6	AAL45696 Human blo
2	1683	100.0	3008	6	ABN95170 Gene #166
3	1683	100.0	3008	7	ACC51204 Human Plk
4	1679.8	99.8	1683	6	AAL45697 Human blo
5	1678.2	99.7	3623	3	AACT6693 Human ORF
6	526.4	31.3	617	3	AAA44763 Human sec
7	368.4	21.9	428	8	ACH21030 Human acu
8	242.6	14.4	397	6	AB199281 Mouse isc
9	120.2	7.1	451	4	ABA58323 Human foe
10	120.2	7.1	451	4	AAI37947 Probe #66
11	120.2	7.1	451	4	AAK32096 Human bon
12	120.2	7.1	451	4	AAK06420 Human bra
13	120.2	7.1	451	4	ABS31791 Human liv
14	120.2	7.1	451	6	ABS06863 Human gen
15	117.6	7.0	2036	3	AAQ63951 Hepatocyt
16	117.6	7.0	2036	6	ABL65242 Lung canc
17	117.6	7.0	2036	6	ABN95114 Gene #161
18	117	7.0	117	4	ABA70919 Human foe
19	117	7.0	117	4	AAI51106 Probe #19
20	117	7.0	117	4	AAK45153 Human bon
21	117	7.0	117	4	AAK13188 Human bra
22	117	7.0	117	4	ABS44824 Human liv
23	117	7.0	117	6	ABS19403 Human gen

24	102.6	6.1	2145	9	ADSE06939	Ade06939	Novel cod
25	96.2	5.7	970	2	AAQ63945	Gene comp	
26	93.2	5.5	2299	7	ABX17688	DNA encod	
27	90.6	5.4	2512	2	AAQ12867	JMI-229 c	
28	89	5.3	2445	9	ADBS2711	Primary r	
29	87.6	5.2	2252	2	AAQ05673	v-PA alph	
30	86.8	5.2	1323	1	AAK70992	Modified	
31	85.2	5.1	1422	1	AAK70991	Modified	
32	84.8	5.0	2137	2	AAQ05675	v-PA_beta	
33	79.6	4.7	2257	2	AAQ05674	v-PA_alpha	
34	78.8	4.7	919	2	AAK90048	Chicken u	
35	78.6	4.7	1689	1	AAK91217	Mutated c	
36	78	4.6	1620	2	AAQ00543	Vampire b	
37	77	4.6	2037	2	AAQ12071	T-PA with	
38	77	4.6	2100	2	AAQ12074	T-PA vari	
39	75.4	4.5	1047	2	AAQ40658	TPA-2 CDN	
40	75.4	4.5	1065	2	AAK87456	mt-PA6-E	
41	75.4	4.5	1085	2	AAK62606	DNA encod	
42	75.4	4.5	1068	2	AAQ53322	Human tru	
43	75.4	4.5	1068	2	AAQ53320	Human tru	
44	75.4	4.5	1068	2	AAQ40683	TPA-5 CDN	
45	75.4	4.5	1068	2	AAQ40657	TPA-1 CDN	

#### ALIGNMENTS

RESULT 1

AAAL45696

ID AAL45696 standard; DNA; 1683 BP.

XX AAL45696;

DT 13-JUN-2002 (first entry)

DE Human blood coagulation factor VII activating protease DNA.

XX Human; blood coagulation factor VII activating protease; FSAP;  
KW single-chain plasminogen activator; bleeding disorder; haematological;  
KW haemostatic; gene; ds.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..1683

FT /tag= a

FT /product= "FSAP"

PN EP1182258-A1.

XX 27-FEB-2002.

XX 05-JUL-2001; 2001EP-00115691.

XX 26-JUL-2000; 2000DE-01036641.

XX 10-OCT-2000; 2000DE-01050040.

XX 21-OCT-2000; 2000DE-01052319.

XX 12-APR-2001; 2001DE-01018706.

XX (AVET ) AVENTIS BEHRING GMBH.

XX Roemisch J, Stoehr H, Feussner A, Lang W, Weimer T, Becker M;

XX Werlich C, Muth-Naumann G;

XX WPI; 2002-270939/32.

XX P-PSDB; AAO17144.

XX New nucleic acid encoding mutant factor 7 activating protease, useful for

XX diagnosis, treatment and prevention of coagulation disorders, also

XX related protein and antibodies.

XX Claim 2; Page 15-16; 27pp; German.

XX



CC The present invention relates to a mutant of the DNA sequence encoding  
 CC the protease (FSAP) that activates blood coagulation factor VII (FVII)  
 CC and single-chain plasminogen activator, where at least one of the base  
 CC changes G to C at nucleotide 1177 and G to A at nucleotide 1601 is  
 CC present. The mutant sequences can be used in the treatment and prevention  
 CC of bleeding disorders associated with inherited or acquired defects of  
 CC blood coagulation factors V, VIII, IX, X, XI, XII, von Willebrand's  
 CC factor and/or with antibodies against any of these proteins. The present  
 CC sequence is the human FSAP coding sequence  
 XX

SQ Sequence 1683 BP; 440 A; 436 C; 437 G; 370 T; 0 U; 0 Other;

Query Match 100.0%; Score 1683; DB 6; Length 1683;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATCTTTGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGCTGGGAAAGACA	60
DB	1	ATCTTTGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGCTGGGAAAGACA	60
QY	61	GCTGTGGGTTCTCCCTGATGCTTTTATTTGGAAGCTGACCCAGACTGGACCCCTGAC	120
DB	61	GCTGTGGGTTCTCCCTGATGCTTTTATTTGGAAGCTGACCCAGACTGGACCCCTGAC	120
QY	121	CAGTATGATTACAGCTACGAGGATTATATTCAGGAAGAGAACACCAAGTAGCACCTTACC	180
DB	121	CAGTATGATTACAGCTACGAGGATTATATTCAGGAAGAGAACACCAAGTAGCACCTTACC	180
QY	181	CATGCTGAGAAATCTGATCTGCTGATACATGAGAGCAAGCTGATCCATGCCAGCCCAAC	240
DB	181	CATGCTGAGAAATCTGATCTGCTGATACATGAGAGCAAGCTGATCCATGCCAGCCCAAC	240
QY	241	CCCTGTGAACACGGTGGGACTGCTGCTCCATGGGAGACCTTACATGCAGCTGCCCTG	300
DB	241	CCCTGTGAACACGGTGGGACTGCTGCTCCATGGGAGACCTTACATGCAGCTGCCCTG	300
QY	301	GCTCTTTCTCTGGGAATAAGTGTGAGAAAGTGAATAATACGTGCAAGACAAACCATGT	360
DB	301	GCTCTTTCTCTGGGAATAAGTGTGAGAAAGTGAATAATACGTGCAAGACAAACCATGT	360
QY	361	GGCGGGGCAATGCTCTCATTACCCAGAGTCTCCCTACTACCGCTGTCTGTAAACAC	420
DB	361	GGCGGGGCAATGCTCTCATTACCCAGAGTCTCCCTACTACCGCTGTCTGTAAACAC	420
QY	421	CTTTACACAGTCCCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	480
DB	421	CTTTACACAGTCCCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	480
QY	481	AATGGGCTACCTGCTCCGCGATACAGGAGATCCAAAGTTCACCTGTGCTGCTGCTGCTG	540
DB	481	AATGGGCTACCTGCTCCGCGATACAGGAGATCCAAAGTTCACCTGTGCTGCTGCTGCTG	540
QY	541	CAGTTCAAGGGGAAATCTGTGAATAGTTCGTGATGACTGCTGATGTTGGCGATGCTAC	600
DB	541	CAGTTCAAGGGGAAATCTGTGAATAGTTCGTGATGACTGCTGATGTTGGCGATGCTAC	600
QY	601	TCCTACCGAGGAAATGAATAGACAGTCAACAGCATGCTGCTGCTTACTGGAATCC	660
DB	601	TCCTACCGAGGAAATGAATAGACAGTCAACAGCATGCTGCTGCTTACTGGAATCC	660
QY	661	CACCTCTCTTTCAGGAGAAATTAACAATGTTTATGGAGATGCTGAAACCCCATGGGATT	720
DB	661	CACCTCTCTTTCAGGAGAAATTAACAATGTTTATGGAGATGCTGAAACCCCATGGGATT	720
QY	721	GGGGAACCAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGCTGCTTATATAA	780
DB	721	GGGGAACCAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGCTGCTTATATAA	780
QY	781	GTTACCAATGACAGGTGAATGGAATGATGCTGATGCTCAGGCTGCTCAGCCAGGAC	840
DB	781	GTTACCAATGACAGGTGAATGGAATGATGCTGATGCTCAGGCTGCTCAGCCAGGAC	840
QY	841	GTTGCTTACCCAGAGGAAAGCCCACTGAGGCATCAACCAAGCTTCGCGGGTTGACTCC	900

DB	841	GTTCCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGCGGGTTGACTCC	900
QY	901	TGTGGAAGACTGAGATAGCAGAGAGAGATCAAGAGAATCTATCGAGCTTTAAGAGC	960
DB	901	TGTGGAAGACTGAGATAGCAGAGAGAGATCAAGAGAATCTATCGAGCTTTAAGAGC	960
QY	961	ACGCGGGCAAGCAACCATCGAGGGCTCCCTCAGTCCCTGCTGCTGCTGCTGCTGCTGCTG	1020
DB	961	ACGCGGGCAAGCAACCATCGAGGGCTCCCTCAGTCCCTGCTGCTGCTGCTGCTGCTGCTG	1020
QY	1021	ATGCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT	1080
DB	1021	ATGCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT	1080
QY	1081	GCCCACTGCACCGCATATAAAACAGACATCTAAAGTGTGCTAGGGACCAAGAGCCTG	1140
DB	1081	GCCCACTGCACCGCATATAAAACAGACATCTAAAGTGTGCTAGGGACCAAGAGCCTG	1140
QY	1141	AAGAAAGAGAAATTCATGAGCAGAGCTTTAGGGTGGAGAGATATTTCAAGTACAGCCAC	1200
DB	1141	AAGAAAGAGAAATTCATGAGCAGAGCTTTAGGGTGGAGAGATATTTCAAGTACAGCCAC	1200
QY	1201	TACAAAGAGAGATGAGATTCCCAACATGATTTGCAATGCTCAAGTTAAAGCCAGTG	1260
DB	1201	TACAAAGAGAGATGAGATTCCCAACATGATTTGCAATGCTCAAGTTAAAGCCAGTG	1260
QY	1261	GATGCTCACTGCTGTAGAAATCCAAATAGCTGAAGACTGTGCTGCTGCTGCTGCTGCTG	1320
DB	1261	GATGCTCACTGCTGTAGAAATCCAAATAGCTGAAGACTGTGCTGCTGCTGCTGCTGCTG	1320
QY	1321	TTTCCCTCTGGGAGTGGCACAATCTCTGGTGGGGTGTACAGAAACAGGAAAGGG	1380
DB	1321	TTTCCCTCTGGGAGTGGCACAATCTCTGGTGGGGTGTACAGAAACAGGAAAGGG	1380
QY	1381	TCCCGCAGCTCTGGGATGCCAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC	1440
DB	1381	TCCCGCAGCTCTGGGATGCCAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC	1440
QY	1441	CAACTCTATGACACATGATTTGATGACAGTATGATCTGTGCGAGAAATCTTCAGAACT	1500
DB	1441	CAACTCTATGACACATGATTTGATGACAGTATGATCTGTGCGAGAAATCTTCAGAACT	1500
QY	1501	GGGCAAGACCTGCCAGGTGCTCTGGAGGCCCCCTGACCTGTGAGAGAGGCGGACCC	1560
DB	1501	GGGCAAGACCTGCCAGGTGCTCTGGAGGCCCCCTGACCTGTGAGAGAGGCGGACCC	1560
QY	1561	TACTACGCTCTATGGGATAGTGAAGTGGGGCTGAGTGTGGGAGAGAGCCAGGGGTCTAC	1620
DB	1561	TACTACGCTCTATGGGATAGTGAAGTGGGGCTGAGTGTGGGAGAGAGCCAGGGGTCTAC	1620
QY	1621	ACCCAGATTCACAAATTCCTGAATGGATCAAGCCCACTCAAAAGTGAAGTGGCTTC	1680
DB	1621	ACCCAGATTCACAAATTCCTGAATGGATCAAGCCCACTCAAAAGTGAAGTGGCTTC	1680
QY	1681	TAA 1683	
DB	1681	TAA 1683	

RESULT 2  
 ABN95170  
 ID ABN95170 standard; DNA; 3008 BP.  
 XX  
 AC ABN95170;  
 XX  
 DT 13-AUG-2002 (first entry)  
 XX  
 DE Gene #1668 used to diagnose liver cancer.  
 XX  
 KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
 KW metastatic liver tumour; cytostatic; expression profile; disease state;  
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.

XX Homo sapiens.  
OS WO200229103-A2.  
FN 11-APR-2002.  
PD 02-OCT-2001; 2001WO-US030589.  
XX 02-OCT-2000; 2000US-0237054P.  
XX (GENE-) GENE LOGIC INC.  
XX  
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
XX WPI; 2002-426119/45.  
DR  
XX  
XX Diagnosing and detecting the progression of liver cancer, hepatocellular  
PT carcinoma or metastatic liver tumor in a patient, involves detecting the  
PT level of expression of two or more genes in a liver tissue sample.  
XX  
XX Claim 1; SEQ ID NO 1668; 298pp; English.  
XX  
XX The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumor in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in AB93303-AB937455 in a  
CC tissue sample. The method of the invention has hepatocytic, and  
CC cytosolic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 3008 BP; 791 A; 837 C; 703 G; 677 T; 0 U; 0 Other;

Query Match 100.0%; Score 1693; DB 6; Length 3008;  
Best Local Similarity 100.0%; Fred. No. 0;  
Matches 1693; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTGTCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60  
DB 97 ATGTTGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 156  
QY 61 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC 120  
DB 157 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC 216  
QY 121 CAGTATGATTACAGCTACGAGATTATTAATCAGGAAGAGACACCAGTAGCACACTTACC 180  
DB 217 CAGTATGATTACAGCTACGAGATTATTAATCAGGAAGAGACACCAGTAGCACACTTACC 276  
QY 181 CATGCTGAGATCTGATGTTGTTACTACTGAGGACCAAGCTGATCCATGCCGCCAAC 240  
DB 277 CATGCTGAGATCTGATGTTGTTACTACTGAGGACCAAGCTGATCCATGCCGCCAAC 336  
QY 241 CCTCTGTAACAGGTTGGGAGCTGCTGCTGCTCATGAGGACCTTTCATGAGCTGCTG 300  
DB 337 CCTCTGTAACAGGTTGGGAGCTGCTGCTGCTCATGAGGACCTTTCATGAGCTGCTG 396  
QY 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACCAACCTATGT 360  
DB 397 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACCAACCTATGT 456  
QY 361 GGCCGGGGCCATGCTCTATTACCCAGAGTCTCTCCTACTACCGTGTGCTGTTAAACAC 420  
DB 457 GGCCGGGGCCATGCTCTATTACCCAGAGTCTCTCCTACTACCGTGTGCTGTTAAACAC 516

QY 421 CTTTACACAGGTCCTCCAGTGTCTCCCAAGTGGTTCCTGTATGTCAGGCCAAACCCCTGCCAG 480  
DB 517 CTTTACACAGGTCCTCCAGTGTCTCCCAAGTGGTTCCTGTATGTCAGGCCAAACCCCTGCCAG 576  
QY 481 AATGGGGCTACCTGCTCCCGGCATTAAGCGAGATCCCAAGTTCACCTGTGSCCTGTCCCGAC 540  
DB 577 AATGGGGCTACCTGCTCCCGGCATTAAGCGAGATCCCAAGTTCACCTGTGSCCTGTCCCGAC 636  
QY 541 CAGTTCAAGGGGAAATTTCTGTGAATAGGTTCTGTAGTCTGTATGTCATGTTGGCGATGCTAC 600  
DB 637 CAGTTCAAGGGGAAATTTCTGTGAATAGGTTCTGTAGTCTGTATGTCATGTTGGCGATGCTAC 696  
QY 601 TCTTACCGAGGGGAAATTAAGTAGGACAGTCAACACAGCATGCGTGTGCTTTTACTGGAACTCC 660  
DB 697 TCTTACCGAGGGGAAATTAAGTAGGACAGTCAACACAGCATGCGTGTGCTTTTACTGGAACTCC 756  
QY 661 CACCTCCTCTTCCAGGAGAAATTAACACATGTTTATGAGGATGCTCAAAACCCATGGATT 720  
DB 757 CACCTCCTCTTCCAGGAGAAATTAACACATGTTTATGAGGATGCTCAAAACCCATGGATT 816  
QY 721 GGGGAACAATTTCTGCAAGAACCCAGATGGGAGGAAAGCCCTGTGCTTTTATAA 780  
DB 817 GGGGAACAATTTCTGCAAGAACCCAGATGGGAGGAAAGCCCTGTGCTTTTATAA 876  
QY 781 GTTACCAATGACAAGGTGAATGGGAATACTGTGATGTTCTCAGCCTGTCTCAGCCCCAGGAC 840  
DB 877 GTTACCAATGACAAGGTGAATGGGAATACTGTGATGTTCTCAGCCTGTCTCAGCCCCAGGAC 936  
QY 841 GTTGCTCCTCCAGAGGAAAGCCCTAGCCATCAACCAAGCTTCCGGGGTTTGAATCC 900  
DB 937 GTTGCTCCTCCAGAGGAAAGCCCTAGCCATCAACCAAGCTTCCGGGGTTTGAATCC 996  
QY 901 TGTGGAAACACTGAGATAGCAGAGAGGAAAGATCAAGAGAAATCTATGGAGGCTTTAAGAGC 960  
DB 997 TGTGGAAACACTGAGATAGCAGAGAGGAAAGATCAAGAGAAATCTATGGAGGCTTTAAGAGC 1056  
QY 961 AGCGGGGCAAGCACCATTGGCAGGCGTCCCTCCAGTCCCTGCTGCTGACCAATCTCC 1020  
DB 1057 AGCGGGGCAAGCACCATTGGCAGGCGTCCCTCCAGTCCCTGCTGCTGACCAATCTCC 1116  
QY 1021 ATGCCCCAGGGCCACTTCTGTGTTGGGCGCTGATCCACCCCTGTCTGGTGTCTCACTGCT 1080  
DB 1117 ATGCCCCAGGGCCACTTCTGTGTTGGGCGCTGATCCACCCCTGTCTGGTGTCTCACTGCT 1176  
QY 1081 GCCCACTGCACGACATAAAGAACCAAGACATCTAAAGTGTGCTAGGGGACAGGACCTG 1140  
DB 1177 GCCCACTGCACGACATAAAGAACCAAGACATCTAAAGTGTGCTAGGGGACAGGACCTG 1236  
QY 1141 AAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAGATATTCAAGTACAGCCAC 1200  
DB 1237 AAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAGATATTCAAGTACAGCCAC 1296  
QY 1201 TACAATGAAGAGATGAGATTCCCAACATGATATTGCAATTCCTCAAGTTAAAGCCAGTG 1260  
DB 1297 TACAATGAAGAGATGAGATTCCCAACATGATATTGCAATTCCTCAAGTTAAAGCCAGTG 1356  
QY 1261 GATGTCATCTGTGCTTAGAATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGGTCC 1320  
DB 1357 GATGTCATCTGTGCTTAGAATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGGTCC 1416  
QY 1321 TTTCCCTCTGGGAGTGGAGTCCACATCTCTGGCTGGGGTGTATTACAGAAACAGGAAAGGG 1380  
DB 1417 TTTCCCTCTGGGAGTGGAGTGGAGTCCACATCTCTGGCTGGGGTGTATTACAGAAACAGGAAAGGG 1476  
QY 1381 TCCCGCAGCTCTGTGATGCCAAAGTCAAGCTGATTGCCAAACACTTTTGTGCAACTCCCGC 1440  
DB 1477 TCCCGCAGCTCTGTGATGCCAAAGTCAAGCTGATTGCCAAACACTTTTGTGCAACTCCCGC 1536  
QY 1441 CAACTCTATGACCAATGATTGATGACAGTATGATGTCGTGAGGAAATCTTTCAGAAACCT 1500  
DB 1537 CAACTCTATGACCAATGATTGATGACAGTATGATGTCGTGAGGAAATCTTTCAGAAACCT 1596  
QY 1501 GGGCAAGACACCTGCCAGGGTGAATCTGGAAGCCGCCCTGACCTGTGAGAAAGGACGGCACC 1560

Db 1597 GGGCAAGACCTGCCAGGGTGACTCTGGAGGCCCTGACCTGTGAGAAGACGGCAC 1656  
 Qy 1561 TACTACGCTCTATGGGATAGTAGCTGGGCTGGAGTGTGGAGAGGCGGAGGCTCTAC 1620  
 Db 1657 TACTACGCTCTATGGGATAGTAGCTGGGCTGGAGTGTGGAGAGGCGGAGGCTCTAC 1716  
 Qy 1621 ACCCAAGTACCAAAATTCCTGAATGGATCAAGGCCACCATCAAAAGTGAAGTGGCTTC 1680  
 Db 1717 ACCCAAGTACCAAAATTCCTGAATGGATCAAGGCCACCATCAAAAGTGAAGTGGCTTC 1776  
 Qy 1681 TAA 1683  
 Db 1777 TAA 1779

RESULT 3  
 ACC51204

ID ACC51204 standard; cDNA; 3008 BP.

XX AC

XX ACC51204;

DT 16-JUN-2003 (first entry)

XX Human Plk-1 related cDNA sequence hmft-0306 SEQ ID NO:89.

XX Human; hepatoblastoma; cancer detection probe; cancer; detection;  
 KW hepatocellular carcinoma; hereditary non-polyposis colorectal cancer;  
 KW desmoid tumour; anaplastic thyroid carcinoma; Wilm's tumour; tumour;  
 KW Plk-1; polo-like Kinase-1; Gene; ss.

XX Homo sapiens.

XX WO2003018807-A1.

XX PD 06-MAR-2003.

XX 26-AUG-2002; 2002WO-JP008580.

XX 24-AUG-2001; 2001JP-00255225.

XX (HISM) HISAMITSU PHARM CO LTD.  
 XX (CHIB-) CHIBA PREFECTURE.

XX Nakagawara A;

XX WPI; 2003-268424/26.

XX Nucleic acid sequences differently expressed between hepatoblastoma and  
 normal liver tissue, are useful for cancer detection and diagnosis.

XX Claim 4; Page 142-144; 180pp; Japanese.

XX The present invention describes nucleic acid sequences (I) having a  
 different degree of expression in hepatoblastoma from their expression in  
 normal liver tissue. ACC51116 to ACC51219 represents specifically claimed  
 examples of (I). Also described: (1) nucleic acids stringently  
 hybridizing to (I); (2) cancer detection probes containing one or more of  
 104 listed sequences (II); see ACC5116 to ACC5119, including the 79 (I),  
 see ACC5116 to ACC51194, or partial sequences derived from them; (3)  
 PCR primers for cancer detection based on sequences (II); (4) marker  
 proteins for cancer detection, encoded by (II); (5) diagnostic reagents  
 for cancer diagnosis, containing (II) or their partial sequences. The  
 nucleic acid sequences are useful in the detection and diagnosis of  
 cancers including liver, colon, breast, kidney, bladder, ovary and  
 thyroid cancer, especially for hepatoblastoma, hepatocellular carcinoma,  
 hereditary non-polyposis colorectal cancer, desmoid tumour, anaplastic  
 thyroid carcinoma and Wilm's tumour. They are also used as markers for  
 predicting the prognosis of these tumours. ACC51220 to ACC51233 represent  
 PCR primers used in the exemplification of the present invention. The  
 nucleic acid sequences given in ACC5116 to ACC51219 are related to human  
 Plk-1 (polo-like kinase-1), which is located on chromosome 16p12

XX

SQ Sequence 3008 BP; 791 A; 837 C; 703 G; 677 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 1683; DB 7; Length 3008;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 ATGTTGCCAGAGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGGTGGAAAGACA 60  
 Db 97 ATGTTGCCAGAGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGGTGGAAAGACA 156  
 Qy 61 GCCTGTGGGTCTCCCTGATGTTTATTGGAAGCTGGACCCAGACCTGGACCCCTGAC 120  
 Db 157 GCCTGTGGGTCTCCCTGATGTTTATTGGAAGCTGGACCCAGACCTGGACCCCTGAC 216  
 Qy 121 CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAGAACACCACTAGTACACTTACC 180  
 Db 217 CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAGAACACCACTAGTACACTTACC 276  
 Qy 181 CATGCTGAGAATCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240  
 Db 277 CATGCTGAGAATCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 336  
 Qy 241 CCCTGTGAACACGGTGGGACTGCTGCTCCATGGGAGACCTTCACTAGCTAGCTGCTG 300  
 Db 337 CCCTGTGAACACGGTGGGACTGCTGCTCCATGGGAGACCTTCACTAGCTAGCTGCTG 396  
 Qy 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTCAAAATACGTGCAAGGACCAACCCATGT 360  
 Db 397 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTCAAAATACGTGCAAGGACCAACCCATGT 456  
 Qy 361 GGCGGGGCCAATGTCCTATACCCAGAGTCCTCCCTACTACCGCTGTGTCTGTAACAC 420  
 Db 457 GGCGGGGCCAATGTCCTATACCCAGAGTCCTCCCTACTACCGCTGTGTCTGTAACAC 516  
 Qy 421 CCTTACACAGCTCCAGCTGCTCCCAAGTGTCTGTTATGAGGAGCAACCCCTGCCAG 480  
 Db 517 CCTTACACAGCTCCAGCTGCTCCCAAGTGTCTGTTATGAGGAGCAACCCCTGCCAG 576  
 Qy 481 AATGGGCTACCTGCTCCCGGCATAAGCGGAGATCAAGTTCACTGTGCTGTCCCGAC 540  
 Db 577 AATGGGCTACCTGCTCCCGGCATAAGCGGAGATCAAGTTCACTGTGCTGTCCCGAC 636  
 Qy 541 CAGTTCAAGGGGAAATTCCTGGAATAGTTCCTGATGACTGCTATGCTTGGCGATGCTAC 600  
 Db 637 CAGTTCAAGGGGAAATTCCTGGAATAGTTCCTGATGACTGCTATGCTTGGCGATGCTAC 696  
 Qy 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCGTGCCTTTACTGGAATCC 660  
 Db 597 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCGTGCCTTTACTGGAATCC 756  
 Qy 561 CACCTCTCTTGAGGAGATTAACAATGTTTATGAGGATGCTGAACCCATGGATT 720  
 Db 757 CACCTCTCTTGAGGAGATTAACAATGTTTATGAGGATGCTGAACCCATGGATT 816  
 Qy 721 GGGGAACACAATTTCTGCAGAAACCCAGATGCGAGGAAAGCCCTGGTCTTTATTA 780  
 Db 817 GGGGAACACAATTTCTGCAGAAACCCAGATGCGAGGAAAGCCCTGGTCTTTATTA 876  
 Qy 781 GTTACCAATGACAAGTGAATGGGAATGCTGATGCTCAGCTGCTCAGCCAGGAC 840  
 Db 877 GTTACCAATGACAAGTGAATGGGAATGCTGATGCTCAGCTGCTCAGCCAGGAC 936  
 Qy 841 GTTCCCTTACCAGAGGAAACCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC 900  
 Db 937 GTTCCCTTACCAGAGGAAACCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC 996  
 Qy 901 TGTGAAAGACTGAGATAGCAGAGAGAGATCAAGAGATCTATGAGGCTTTAGAGC 960  
 Db 997 TGTGAAAGACTGAGATAGCAGAGAGAGATCAAGAGATCTATGAGGCTTTAGAGC 1056  
 Qy 961 ACGCCGGGCAAGCACCCATGGAGGCTCCCTCAGTCTCGCTGCTGCTGACCATCTCC 1020  
 Db 1057 ACGCCGGGCAAGCACCCATGGAGGCTCCCTCAGTCTCGCTGCTGCTGACCATCTCC 1116

QY	1021	ATGCCCCAGGGCCACTTCTGTGTGGTGGGGCGCTGATCCACCCCTGTGCGGTGTCTACTGCT	1080
Db	1117	ATGCCCCAGGGCCACTTCTGTGTGGTGGGGCGCTGATCCACCCCTGTGCGGTGTCTACTGCT	1176
QY	1081	GCCCACTGACCGACATATAAACCAGACATCTAAAGGTGGTGTCTAGGGACGACGACCTG	1140
Db	1177	GCCCACTGACCGACATATAAACCAGACATCTAAAGGTGGTGTCTAGGGACGACGACCTG	1236
QY	1141	AGAAAAGAAATTTTCATGACGACAGACTTTAGGGTGGAGAAAGATATTCAAGTACAGCCAC	1200
Db	1237	AGAAAAGAAATTTTCATGACGACAGACTTTAGGGTGGAGAAAGATATTCAAGTACAGCCAC	1296
QY	1201	TACAAATGAAGAGATGAGATTTCCCAACATGATATTGCAATTTGCTCAAGTTAAAGCCAGTG	1260
Db	1297	TACAAATGAAGAGATGAGATTTCCCAACATGATATTGCAATTTGCTCAAGTTAAAGCCAGTG	1356
QY	1261	GATGGTCACTGTGCTCTAGAAATCCAAATACGTCAAGACTGTGTGCTTTGCTGATGGGTCC	1320
Db	1357	GATGGTCACTGTGCTCTAGAAATCCAAATACGTCAAGACTGTGTGCTTTGCTGATGGGTCC	1416
QY	1321	TTTCCCTCTGGGAGTGAGTGCCCACTCTCTGGGTGGGGTGTTCACAGAAACAGAAAGGG	1380
Db	1417	TTTCCCTCTGGGAGTGAGTGCCCACTCTCTGGGTGGGGTGTTCACAGAAACAGAAAGGG	1476
QY	1381	TCCGGCCAGCTCTTGGATGCAAGTCAAGCTGATTTGCCCAACACTTTGTGCAACTCCCCC	1440
Db	1477	TCCGGCCAGCTCTTGGATGCAAGTCAAGCTGATTTGCCCAACACTTTGTGCAACTCCCCC	1536
QY	1441	CAACTCTATGACCACATGATTGATGACAGTATGATCTGTGCAGGAAATCTTCAGAAACCT	1500
Db	1537	CAACTCTATGACCACATGATTGATGACAGTATGATCTGTGCAGGAAATCTTCAGAAACCT	1596
QY	1501	GGGCAAGACACTGCGCAGGTGACTCTGAGGGCCCCCTGACCTGTGAGAAGACGGCACC	1560
Db	1597	GGGCAAGACACTGCGCAGGTGACTCTGAGGGCCCCCTGACCTGTGAGAAGACGGCACC	1656
QY	1561	TACTACTCTTATGGGATAGTCAAGCTGGGGCTGGAGTGTGGAGAGAGCCACAGGGTCTAC	1620
Db	1657	TACTACTCTTATGGGATAGTCAAGCTGGGGCTGGAGTGTGGAGAGAGCCACAGGGTCTAC	1716
QY	1621	ACCAAAGTTACCAAAATTCCTGAAATTGGATCAAAGCCCAACCTCAAAAGCTGAAAGTGGCTTC	1680
Db	1717	ACCAAAGTTACCAAAATTCCTGAAATTGGATCAAAGCCCAACCTCAAAAGCTGAAAGTGGCTTC	1776
QY	1681	TAA	1683
Db	1777	TAA	1779

## RESULT 4

AA145697			
ID	AA145697 standard; DNA; 1683 BP.		
XX			
AC	AA145697;		
XX			
DT	13-JUN-2002 (first entry)		
XX			
DE	Human blood coagulation factor VII activating protease mutant DNA.		
XX			
XX	Human; blood coagulation factor VII activating protease; PSAP;		
KW	single-chain plasminogen activator; bleeding disorder; haematological;		
KW	haemostatic; mutant; gene; ds.		
XX			
OS	Homo sapiens.		
OS	Synthetic.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..1683	
FT		/*tag= a	
FT		/product= "mutant PSAP"	
XX			
PN	EP1182258-A1.		

XX	27-FEB-2002.	
PD		
XX	05-JUL-2001; 2001EP-00115691.	
XX		
XX	26-JUL-2000; 2000DE-01035641.	
PR	10-OCT-2000; 2000DE-01050040.	
PR	21-OCT-2000; 2000DE-01052319.	
PR	12-APR-2001; 2001DE-01018706.	
XX		
XX	(AVET ) AVENTIS BEHRING GMBH.	
PA		
XX	Roemisch J, Stoehr H, Feussner A, Lang W, Weimer T, Becker M;	
PI	Nerlich C, Muth-Naumann G;	
PI		
XX	WPI; 2002-270939/32.	
DR	P-PSDB; AAO17145.	
DR		
XX	New nucleic acid encoding mutant factor 7 activating protease, useful for	
PT	diagnosis, treatment and prevention of coagulation disorders, also	
PT	related protein and antibodies.	
XX		
XX	Disclosure; Page 16-17; 27pp; German.	
FS		
XX	The present invention relates to a mutant of the DNA sequence encoding	
CC	the protease (FSAP) that activates blood coagulation factor VII (FVII)	
CC	and single-chain plasminogen activator, where at least one of the base	
CC	changes G to C at nucleotide 1177 and G to A at nucleotide 1601 is	
CC	present. The mutant sequences can be used in the treatment and prevention	
CC	of bleeding disorders associated with inherited or acquired defects of	
CC	blood coagulation factors V, VIII, IX, X, XI, XII, von Willebrand's	
CC	factor and/or with antibodies against any of these proteins. The present	
CC	sequence is the mutant human FSAP coding sequence	
XX		
SO	Sequence 1683 BP; 441 A; 437 C; 435 G; 370 T; 0 U; 0 Other;	

Query Match	99.8%	Score 1679.8	DB 6	Length 1683
Best Local Similarity	99.9%	Pred. No. 0		
Matches 1681	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY	1	ATGTTTGCCAGGATGTCTGATCTCCATGTCTTCTGCTTAATGGCTCTGGTGGGAAGACA	60	
DB	1	ATGTTTGCCAGGATGTCTGATCTCCATGTCTTCTGCTTAATGGCTCTGGTGGGAAGACA	60	
QY	61	GCCTGTGGGTTCTCCCTGTAGTCTTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC	120	
DB	61	GCCTGTGGGTTCTCCCTGTAGTCTTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC	120	
QY	121	CAGTATGATTACAGCTACGAGGATTATAATCAGGAAGAGAACCCAGTAGCACATTACC	180	
DB	121	CAGTATGATTACAGCTACGAGGATTATAATCAGGAAGAGAACCCAGTAGCACATTACC	180	
QY	181	CATGCTGAGAAATCTTGACTGGTACTACACTGAGACCAAGCTGATCCATGCCAGGCCAAC	240	
DB	181	CATGCTGAGAAATCTTGACTGGTACTACACTGAGACCAAGCTGATCCATGCCAGGCCAAC	240	
QY	241	CCCTGTGAACCGGTGGGGACTGCTCTGTCATGGGAGCACCTTTCACATGCAGTGCCTG	300	
DB	241	CCCTGTGAACCGGTGGGGACTGCTCTGTCATGGGAGCACCTTTCACATGCAGTGCCTG	300	
QY	301	GCTCCTTTCTCTGGGAATAAGTGTCAAGAGTGCAGAAATACGTGCAGAGGACAAACCCATGT	360	
DB	301	GCTCCTTTCTCTGGGAATAAGTGTCAAGAGTGCAGAAATACGTGCAGAGGACAAACCCATGT	360	
QY	361	GGCCGGGGCCCAATGTCTCATTTACCCAGAGTCCCTCCCTACTACCCTGTGTCTGTAACAC	420	
DB	361	GGCCGGGGCCCAATGTCTCATTTACCCAGAGTCCCTCCCTACTACCCTGTGTCTGTAACAC	420	
QY	421	CCTTACACAGTCCCGAGTCTCCCAAGTGGTTCCTGTATCAGGCCCAAAACCCCTGCCAG	480	
DB	421	CCTTACACAGTCCCGAGTCTCCCAAGTGGTTCCTGTATCAGGCCCAAAACCCCTGCCAG	480	
QY	481	AATGGGGGTACTCCTGCTCCCGGCATGAAGCGGAGATCCAAAGTTTCACTGTGCTGTCCCGAC	540	



CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive  
 XX  
 SQ Sequence 3623 BP; 848 A; 1005 C; 930 G; 840 T; 0 U; 0 Other

Sequence 3623 BP; 848 A; 1005 C; 930 G; 840 T; 0 U; 0 Other;

Query Match 99.7%; Score 1678.2; DB 3; Length 3623;

Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1680; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATGTTTGCAGGATGTCGATCTCCATGTTCTGCTGTTAAATGGCTCTGCTGGGAAGACA 60

Db 97 ATGTTTGCAGGATGTCGATCTCCATGTTCTGCTGTTAAATGGCTCTGCTGGGAAGACA 156

Qy 61 GCCTGTGGGTTCTCCCTGATGTCCTTATTGGAAAGCCTGGACCCAGACTGGACCCCTGAC 120

Dp 157 GCCTGTGGGTTCTCCCTGATGTCCTTATTGGAAAGCCTGGACCCAGACTGGACCCCTGAC 216

Qy	121	CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAAACCCAGTAGCACACTTACC	180
Dh	217	CAGCTATGATTACAGCTACGAGGATTATATCAGGAAGAAACCCAGTAGCACACTTACC	276

[illegible]

Qy 241 CCTGTGACACGGTGGGACTGCCTCGTCATGGGAGCACCTTCACATGCAGCTGCCTG 300

Qy 301 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAAAATACGTCCAGGACAACCCATCT 360

361 GGCGGGGCAATGTCCTATTACCAAGTCTCTCCCTACTACCGCTGTCTGTAAAC 420

437 GGCCTGGGCGCCAAATGCTCATATCCCGAGATCCCTCCCTACATATACCCCTGATGCTGATTAACAC 319

421 CCTTACACAGGTCGCCAGCTGCTCCCAAGTGGTTCCTGTATGACGGCCAAACCCCTGCCAG 480

QY

DB	517	CCCTACACAGGTCCCAAGCTGCTCCCAAGGTTCTCTGTAAGCAGGCCAAACCTCTGCCAG	518
Qy	481	AAATGGGGTACCTGCTCCCGGCATAAGCGGAGATCCAAAGTTCACTGTGCCTGTCCCGAC	540

DB	577	AAATGGGGGCTACCTGCTCCCGGCAATAAGCGAGATCCAAAGTTTCACTGTGCTCTGTCCCGAC	636
QY	541	CAGTTCAAGGGGGAAATTCGTGGAATAGTTTCTGATGACTGCTATGTTGCGCGATGGCTAC	600

DB	637	CAGTTCAAGGGGAAATTCGTGAAATAGGTTCTGATGACTGCTATGTTGGCGATGGCTAC	690
QY	601	TCTTACCGGGGAAATGATATGACACAGTCAAACGACGTCGTCGCTTACTTGAACATCC	660

Db	697	TCATTACCGGGAATAATGTAATGACAGCTCAACCAGCATGCGTGCCTTTACTGGAAC	750
Qy	661	CACCTCCTTTGCAGGAGAAATTACAACATGTTTATGGAGATGCTGAAACCATGGGATT	720

DB	757	CACCTCCCTTCGACGAGAAATTACAACATGTTTATGGAGGATCTGTAACCCATGGGATT	815
QY	721	GGGGAACACAAATTTCTGCAGAAACCCAGATCGCGACGAAAAGCCCTGTCCTTTATTTAA	780

Accession	Sequence	Length
Db	GGGGAACAAATTTCTGCAGAAACCCAGATGCGACGAAAGCCCTGGTCTTTATTTAA	875
QV	GTTCACCAATGACAAGGTGAAATGGGAATCTGTGATGTCTCAGCCTGCTCAGGCCCAGGAC	840

Db	877	GTTCACCAATGACAAAGGTGAATGGAATCTGTGATGTCTCAGCTGCTCAGCCACGAG	936
Ov	841	GTTCGCTACCCAGAGGAAAGCCCCCACTGAGCCATCAACCAAGCTTCGCGGGTTTGCATCC	900



thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antidiarrhetic; vulnerary; antiparkinsonian; antifuror; osteopathic; neuroprotective; nootropic; antiparasitic; cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine; autoimmune disorder; multiple sclerosis; allergic condition; insulin dependent diabetes; asthma; myeloid cell osteoclast; ulcer; lymphoid cell deficiency; burn; osteoporosis; osteoarthritis; central nervous system disorder; Alzheimer's disease; stroke; Parkinson's disease; Huntington's disease; coagulation disorder; haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour; infection; depression; psoriasis; ss

Homo sapiens.

WO200021991-A1.

20-APR-2000.

15-OCT-1999: 99WO-US024206.

15-OCT-1998: 98US-0104436P.

(GEMY ) GENETICS INST INC.

JACOBS K MCCOY TM LAYALLIE ER COLLINS-RACIE IA EVANS C

Merberg D, Treacy M

Isolated polynucleotides, and encoded proteins, comprising secreted expressed sequence tags (sESTs), useful for treating various disorders such as autoimmune, infectious, and central nervous system disorders.

Claim 1: Page 549; 803pp; English.

AAA43426 to AAA45926 represent specifically claimed secreted expressed sequence tags (sESTs), isolated from human, mouse, chicken and rat tissue sources. The sESTs can have a range of activities depending on the tissues they were isolated from. The activities include: chemotactic; proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antiasthmatic; vulnery; antitumor; osteopathic; neuroprotective; nootropic; antiparkinsonian; antipsoriatic; cerebroprotective; anticonvulsant; and antidepressant. The sESTs can be used for gene therapy and in vaccines. The sESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the sESTs. Proteins encoded by the sESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's disease), tumours, bacterial, fungal or viral infections, depression and psoriasis. AAA45926 to AAA45931 represent linker variants which are given in the exemplification of the present invention.

Sequence 617 BP; 165 A; 169 C; 148 G; 135 T; 0 U; 0 Other;

every Match	31.3%;	Score 526.4;	DB 3;	Length 617;
ry Local Similarity	99.8%;	Pred. No. 3.1e-140;		
riches 527:	Conservative	0;	Mismatches 1;	Indels 0;
Gaps				

1 ATGTTTGCAGGATCTCGATCTCCATGTTCTGCTGTTAAATGGCTCTGGTGGGAAGACA 60  
84 ATGTTTGCAGGATCTCGATCTCCATGTTCTGCTGTTAAATGGCTCTGGTGGGAAGACA 143  
61 GCCTGTGGGTTCTCCCTGATGTCCTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGCAC 120  
144 GCCTGTGGGTTCTCCCTGATGTCCTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGCAC 203  
121 CAGTATGATTACAGCTACGAGGATTATTAATCAGGAAGAGAACCCAGTAGACACTTACC 180

204	DB	CAGTATGATTTACAGCTACGAGGATTTATATCAGAAAGAGAACCCAGTAGCACATTACC	263
181	QY	CATGCTGAGAATCCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCGAGCCCAAC	240
264	DB	CAGCGTGAGAAATCCTGACTGCTACTACATCTGAGGACCAAGCTGATCCATGCGAGCCCAAC	323
241	QY	CCCTGTGAAACAGGTGGGACCTGCTGCTCCATGGGAGCACCTTACATGCGAGCTGCCCTG	300
324	DB	CCCTGTGAAACAGGTGGGACCTGCTGCTCCATGGGAGCACCTTACATGCGAGCTGCCCTG	383
301	QY	GCTCCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAACCCATGT	360
384	DB	GCTCCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAACCCATGT	443
361	QY	GGCGGGGCGCAATGTCTCATTTACCCAGAGTCCCTCCCTACTACCGCTGTGCTGTGTAACAC	420
444	DB	GGCGGGGCGCAATGTCTCATTTACCCAGAGTCCCTCCCTACTACCGCTGTGCTGTGTAACAC	503
421	QY	CCTTTACACAGTCCGAGCTGTGCTCCCAAGTGGTTCCTGTATGCAAGGCCAAACCCCTGCCAG	480
504	DB	CCTTTACACAGTCCGAGCTGTGCTCCCAAGTGGTTCCTGTATGCAAGGCCAAACCCCTGCCAG	563
481	QY	AATGGGGTCACTGCTCCGGCATTAAGCGGAGATCCCAAGTTCACTGT	528
564	DB	AATGGGGTCACTGCTCCGGCATTAAGCGGAGATCCCAAGTTCACTGT	611

## RESULT 7

ACH21030  
ID ACH

XX AC ACH21030;

XX  
DT 13-OCT-20

XX Human adult liver cDNA #64

XX Human; ss: sequencing by hybridization

**KW** genome mapping; biodiversity; genetic disorder.

XX  
OS Homo sapiens.

XX  
PN US2003073623-

XX  
PD 17-APR-2003.XX  
PF  
30-JUL-2001;

XX  
PR 30-JUL-2001: 2001US-00918995.

XX  
PA (DRMA/) DRMANAC R T.

PA (LABA/) LABAT I.  
PA (STAC/) STACHE-CRAIN

PA (DICK//) DICKSON M C.  
PA (JONE//) JONES L W

XXI  
PI  
Drmanac RT. Labat T  
(SOREN) J. SOREN E. R.

II  
 XX  
 DP  
 DIMENSIONALITY, 2003-615964/58  
 WPI: 2003-615964/58

DR XX  
XX  
PT  
New polynucleotide s  
NEI, 2003-013304/200

new polynucleotide sequences obtained from patients F1 as hybridization probes, as oligomers for PCR, for chromosome PT as hybridization probes, as oligomers for PCR, for chromosome PT as hybridization probes, as oligomers for PCR, or in general PT

PT antisense DNA or RNA  
yy

PS Claim 1; SEQ ID NO 8242; 44pp; English.

XX  
CC The invention relates to an isolated po  
CC 20043 DNA sequence appearing as ACH1

38043 cDNA sequences, appearing as 38147/35210, these sequences determined by the technique of SBH (sequencing by hybridisation). Also included is a purified polypeptide comprising a sequence corresponding to a reading frame of the novel polynucleotide. The nucleic acid sequences



are useful in diagnostics as expressed sequence tags (EST) for  
 CC identifying expressed genes or for physical mapping of the human genome,  
 CC in forensics, in assessing biodiversity, or in identifying mutations  
 CC responsible for genetic disorders and other traits. The nucleotide  
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
 CC for chromosome and gene mapping, in the recombinant production of  
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
 CC is useful for generating antibodies specific for it. The present sequence  
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?docID=20030073623  
 SQ  
 Query Match 21.9%; Score 368.4; DB 8; Length 428;  
 Best Local Similarity 99.7%; Pred. No. 5.2e-95;  
 Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATGTTTGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAAGACA 60  
 Db 59 ATGTTTGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAAGACA 118  
 QY 61 GCTCTGGGTTCTCCCTGATGCTTTATGGAAAGCTTGGACCCAGACTGGACCCCTGAC 120  
 Db 119 GCTCTGGGTTCTCCCTGATGCTTTATGGAAAGCTTGGACCCAGACTGGACCCCTGAC 178  
 QY 121 CAGTATGATTACAGCTACGAGGATTATATCAGGAGAGAACACACAGTAGCACACTTACC 180  
 Db 179 CAGTATGATTACAGCTACGAGGATTATATCAGGAGAGAACACACAGTAGCACACTTACC 238  
 QY 181 CATGCTGAGAACTCTGACTGTGTAATCACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240  
 Db 239 CATGCTGAGAACTCTGACTGTGTAATCACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 298  
 QY 241 CCTGTGAAACAGCTGGGACTGCTCGTCCATGGGAGCACCTTCACATGCAGCTGCCTG 300  
 Db 299 CCTGTGAAACAGCTGGGACTGCTCGTCCATGGGAGCACCTTCACATGCAGCTGCCTG 358  
 QY 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAGAAATACGTGCAAGAGCAACCCATGT 360  
 Db 359 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAGAAATACGTGCAAGAGCAACCCATGT 418  
 QY 361 GGCGGGGGCC 370  
 Db 419 GGCGGGGGCC 428  
 RESULT 8  
 ID ABI9281  
 XX ABI9281 standard; cDNA; 397 BP.  
 AC ABI9281;  
 XX  
 DT 07-MAR-2002 (first entry)  
 XX  
 DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:123.  
 XX  
 KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.  
 XX  
 OS Mus musculus.  
 XX  
 EN WO200188188-A2.  
 XX  
 PD 22-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001WO-JP004192.  
 XX  
 PR 18-MAY-2000; 2000JP-00145977.  
 XX  
 PA (UTNI-) UNIV NIHOON SCHOOL JURIDICAL PERSON.  
 XX

Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
 WPI; 2002-034733/04.  
 XX  
 XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or by  
 PT determining the expression profile of a gene group comprising these  
 PT genes.  
 XX  
 PS Claim 2; Page 351-352; 2690pp; English.  
 XX  
 CC The present invention describes a method for examining ischaemic  
 CC conditions, comprising measuring the expression levels of particular  
 CC genes (1) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes selected from (1). The method  
 CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
 CC expression levels of particular genes (ABI9281 to ABI9914, encoding the  
 CC protein sequences in ABB57020 to ABB57374) or by determining the  
 CC expression profile of a gene group comprising these genes. The expression  
 CC levels or expression profiles produced by these genes are used as an  
 CC indicator when screening for ischaemic condition-improving drugs or  
 CC therapeutics for ischaemic diseases. ABI9913 and ABI9914 represent PCR  
 CC primers for a mouse ischaemic condition related sequence, which are used  
 CC in the exemplification of the present invention  
 XX  
 SQ Sequence 397 BP; 103 A; 105 C; 111 G; 78 T; 0 U; 0 Other;  
 Query Match 14.4%; Score 242.6; DB 6; Length 397;  
 Best Local Similarity 84.7%; Pred. No. 6e-59;  
 Matches 272; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 QY 1363 ACAGAAACAGAAAGGTCCTCCAGCTCTCTGGATGCCAAAGCTCAAGCTGATTGGCAAC 1422  
 Db 13 ACAGAAACAGAAAGGTCCTCCAGCTCTCTGGATGCCAAAGCTCAAGCTGATTGGCAAC 72  
 QY 1423 ACTTTGTGCAACTCCGCCAACTCTATGACCAATGATGATGACAGTATGATCTGTGCA 1482  
 Db 73 CTTTGTGCAACTCCGCCAACTCTATGACCAATGATGATGATGATGATGATGATGCGG 132  
 QY 1483 GGAAATCTTCAGAAACCTGGGCAAGACACCTGCCAGGTGACCTCTGGAGCCGCCCTGACC 1542  
 Db 133 GGGAACTCTCAGAAAGCCCGGATCAGACCTTCCAGGGTACCTCGGGGGGCCCTCTAAC 192  
 QY 1543 TGTGAGAAAGGACGCGCACTTACTAGCTCTATGGATAGTAGCTGGGCGCTGGAGTGTGGG 1602  
 Db 193 TGTGAGAAAGGATGGAATCTTACTAGCTCTACGGGATGTAAGCTGGGCGCGGAAATGTGGG 252  
 QY 1603 AAGAGGCGAGGGTCTACACCCCAAGTTACCAATTCCTGAATTGGATCAAGCCACCATC 1662  
 Db 253 AAGAGGCGAGGGTCTACACTCAAGTCAACCAAGTTCCTGAATTGGATCAAGCCACCATC 312  
 QY 1663 AAAAGTGAAAGTGGGCTTCTAA 1683  
 Db 313 CACAGGAGGCTGGCCTCTGA 333  
 RESULT 9  
 ID ABA58323/c  
 ID ABA58323 standard; DNA; 451 BP.  
 XX  
 XX ABA58323;  
 AC ABA58323;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE Human foetal liver single exon nucleic acid probe #6628.  
 XX  
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 EN WO200157277-A2.  
 XX

PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632356.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 XX WPI; 2001-488897/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 XX gene expression in human placenta.  
 XX  
 PS Claim 25; SEQ ID NO 6633; 654pp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes (SENP).  
 CC The present sequence is one such probe. The probes are useful for  
 CC producing a microarray for predicting, measuring and displaying gene  
 CC expression in samples derived from human placenta. The probes are useful  
 CC for antenatal diagnosis of human genetic disorders  
 XX  
 SQ Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;  
  
 Query Match 7.1%; Score 120.2; DB 4; Length 451;  
 Best Local Similarity 97.6%; Pred. No. 8.1e-24;  
 Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 324 TCAGAAGTGC AAAATACGTGCAAGGACAAACCCATGTGGCCGGGCCAATGTCCTCATTTAC 383  
 Db 451 TCCTACAGTGC AAAATACGTGCAAGGACAAACCCATGTGGCCGGGCCAATGTCCTCATTTAC 392  
  
 QY 384 CCAGAGTCTCTCCCTACTACCGGTGTGTCTGTATAACACCTTACACAGGTCCAGCTGCTC 443  
 Db 391 CCAGAGTCTCTCCCTACTACCGGTGTGTCTGTATAACACCTTACACAGGTCCAGCTGCTC 332  
  
 QY 444 CCAAG 448  
 Db 331 CCAAG 327  
  
 RESULT 11  
 AAK32096/c  
 ID AAK32096 standard; DNA; 451 BP.  
 XX  
 AC AAK32096;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human bone marrow expressed single exon probe SEQ ID NO: 6653.  
 XX  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157276-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000868.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human bone marrow.  
XX  
XX Example 4; SEQ ID NO 6653; 658pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention  
XX  
XX Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;  
SQ  
Query Match 7.1%; Score 120.2; DB 4; Length 451;  
Best Local Similarity 97.6%; Pred. No. 8.1e-24;  
Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 324 TCAGAAAGTGCAAAATACGTGCAAGGACACCCCATGTGGCCGGGCCAATGTCTCATTAC 383  
Dd 451 TCTACAGTGCAAAATACGTGCAAGGACACCCCATGTGGCCGGGCCAATGTCTCATTAC 392  
QY 384 CCAGAGTCTCCCTACTACCGCTGTGTGTGTAAACACCCCTTACACAGGTCCAGCTGCTC 443  
Dd 391 CCAGAGTCTCCCTACTACCGCTGTGTGTGTAAACACCCCTTACACAGGTCCAGCTGCTC 332  
QY 444 CCAAG 448  
Dd 331 CCAAG 327

RESULT 12  
AAK06420/c  
ID AAK06420 standard; DNA; 451 BP.  
XX  
XX AAK06420;  
AC  
XX 05-NOV-2001 (first entry)  
DT  
XX Human brain expressed single exon probe SEQ ID NO: 6411.  
DE  
XX Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;  
KW ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200157275-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 30-JAN-2001; 2001WO-US0000667.  
PF  
XX 04-FEB-2000; 2000US-0180312P.  
PR  
XX 26-MAY-2000; 2000US-0207456P.  
PR  
XX 30-JUN-2000; 2000US-00608408.  
PR  
XX 03-AUG-2000; 2000US-00632366.  
PR  
XX 21-SEP-2000; 2000US-0234687P.  
PR  
XX 27-SEP-2000; 2000US-0236359P.  
PR  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human

PT brains.  
XX  
XX Example 4; SEQ ID NO 6411; 650pp + Sequence Listing; English.  
PS  
XX  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention  
XX  
XX Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;  
SQ  
Query Match 7.1%; Score 120.2; DB 4; Length 451;  
Best Local Similarity 97.6%; Pred. No. 8.1e-24;  
Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 324 TCAGAAAGTGCAAAATACGTGCAAGGACACCCCATGTGGCCGGGCCAATGTCTCATTAC 383  
Dd 451 TCTACAGTGCAAAATACGTGCAAGGACACCCCATGTGGCCGGGCCAATGTCTCATTAC 392  
QY 384 CCAGAGTCTCCCTACTACCGCTGTGTGTGTAAACACCCCTTACACAGGTCCAGCTGCTC 443  
Dd 391 CCAGAGTCTCCCTACTACCGCTGTGTGTGTAAACACCCCTTACACAGGTCCAGCTGCTC 332  
QY 444 CCAAG 448  
Dd 331 CCAAG 327

RESULT 13  
ABS31791/c  
ID ABS31791 standard; DNA; 451 BP.  
XX  
XX ABS31791;  
AC  
XX 25-FEB-2003 (first entry)  
DT  
XX Human liver single exon probe, SEQ ID No 6781.  
Dd  
XX Human; single exon nucleic acid probe; liver; cirrhosis;  
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;  
KW coronary heart disease; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200157273-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 30-JAN-2001; 2001WO-US0000664.  
PF  
XX 04-FEB-2000; 2000US-0180312P.  
PR  
XX 26-MAY-2000; 2000US-0207456P.  
PR  
XX 30-JUN-2000; 2000US-00608408.  
PR  
XX 03-AUG-2000; 2000US-00632366.  
PR  
XX 21-SEP-2000; 2000US-0234687P.  
PR  
XX 27-SEP-2000; 2000US-0236359P.  
PR  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX WPI; 2001-488898/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human adult liver.  
XX  
XX Claim 1; SEQ ID NO 6781; 658pp; English.  
PS  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for

measuring human gene expression in a sample derived from human adult liver, comprising one of 13109 defined nucleotide sequences given in the specification (or complements/fragments). The probe hybridises at high stringency to a nucleic acid molecule expressed in the human adult liver. (1) may be used for predicting, measuring and displaying gene expression in samples derived from human adult liver. The genes identified may be involved in genetic liver diseases such as cirrhosis, hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is associated with coronary heart disease. ABS25011-ABS51005 represent human liver single exon nucleic acid probes of the invention. Note: The sequence information for this patent does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;

Query Match 7.1%; Score 120.2; DB 4; Length 451;  
Best Local Similarity 97.6%; Pred. No. 8.1e-24;  
Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 324 TCAGAAAGTGCAGGACCAACCCATGTGCGCGGCGCAATGCTCATATAC 383  
DB 451 TCCTACAGTGCAAAATACGTGCAAGGACCAACCCATGTGCGCGGCGCAATGCTCATATAC 392

QY 384 CCAGAGTCCCTCTACTACCGCTGTGCTGTAACACCCCTTACACAGGTCCTCAGGTGCTC 443  
DB 391 CCAGAGTCCCTCTACTACCGCTGTGCTGTAACACCCCTTACACAGGTCCTCAGGTGCTC 332

QY 444 CCAAG 448  
DB 331 CCAAG 327

RESULT 14  
ABS06863/C  
ID ABS06863 standard; DNA; 451 BP.  
AC ABS06863;  
XX  
XX 19-AUG-2002 (first entry)  
XX Human genome-derived single exon probe from lung SEQ ID NO 6854.  
XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease.  
XX Homo sapiens.  
OS  
XX  
XX WO200186003-A2.  
XX  
XX 15-NOV-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000665.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00652366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX

WIPI; 2002-114183/15.  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.  
XX  
XX Claim 1; SEQ ID NO 6854; 634pp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived of  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC; the novel set of probes which hybridise at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a single exon probe of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;  
Query Match 7.1%; Score 120.2; DB 6; Length 451;  
Best Local Similarity 97.6%; Pred. No. 8.1e-24;  
Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 324 TCAGAAAGTGCAGGACCAACCCATGTGCGCGGCGCAATGCTCATATAC 383  
DB 451 TCCTACAGTGCAAAATACGTGCAAGGACCAACCCATGTGCGCGGCGCAATGCTCATATAC 392

QY 384 CCAGAGTCCCTCTACTACCGCTGTGCTGTAACACCCCTTACACAGGTCCTCAGGTGCTC 443  
DB 391 CCAGAGTCCCTCTACTACCGCTGTGCTGTAACACCCCTTACACAGGTCCTCAGGTGCTC 332

QY 444 CCAAG 448  
DB 331 CCAAG 327

RESULT 15  
AAQ63951  
ID AAQ63951 standard; cDNA; 2033 BP.  
XX  
XX AAQ63951;  
XX AC  
XX 25-MAR-2003 (revised)  
DT 06-JAN-1995 (first entry)  
DT

XX Hepatocyte growth factor converting protease coding sequence.  
DE  
XX Hepatocyte growth factor; protease; cleavage; active; inactive;  
KW precursor; ds.  
XX Homo sapiens.  
OS  
XX Location/Qualifiers  
FH 1.1968  
FT /\*tag= a  
FT /product= "Hepatocyte growth factor converting protease."  
XX  
XX EP596524-A2.  
PN  
XX 11-MAY-1994.  
PD  
XX 05-NOV-1993; 93EP-00117988.  
PF  
XX 05-NOV-1992; 92JP-00286133.  
PR  
XX 20-NOV-1992; 92JP-00312234.  
PR  
XX 20-NOV-1992; 92JP-00312242.  
PR  
XX (MITU ) MITSUBISHI KASEI CORP.  
PA  
XX Shimomura T, Yamada K, Morimoto Y, Kitamura N, Miyazawa K;  
PI WPI; 1994-152921/19.  
XX P-PSDB; AAR53962.  
DR  
XX Hepatocyte growth factor converting protease and precursor and gene  
XX encoding them - for producing active two chain HGF from inactive single  
XX chain HGF.  
XX  
XX Claim 15; Page 26-27; 30pp; English.  
XX  
XX The polypeptide encoded by this sequence has protease activity and is  
XX capable of converting inactive single chain hepatocyte growth factor  
XX (HGF) into active two chain HGF by cleavage at a specific site. (Updated  
XX on 25-MAR-2003 to correct PN field.)  
XX  
XX Sequence 2033 BP; 339 A; 739 C; 623 G; 332 T; 0 U; 0 Other;  
SQ  
Query Match 7.0%; Score 117.6; DB 2; Length 2033;  
Best Local Similarity 49.8%; Pred. No. 9.2e-23;  
Matches 607; Conservative 0; Mismatches 559; Indels 54; Gaps 10;  
QY 469 AACCCCTGCCAGATGGGGTACTCTGCTCCGCGCATAGGGGATCCAGTTCACCTGT 528  
Db 742 AGCCCTTGCTGAACGGGGACCTGCCACCTGATCTGGCCACCGGACCCCGTGT 801  
QY 529 GCCTGTCCCGACAGTTCAAGGGGAAATCTGTGAATAGTGTCTGATGA---CTGCTAT 585  
Db 802 GCCTGCCACAGGCTTCGTGACGGCTCTGCAACATCGAGCTGTATGAGCGCTGCTTC 861  
QY 586 GTTGGCGATGGCTACTCTTACCGAGGGGAAATGAATAGGACATCAACAGCATGGCTGC 645  
Db 862 TTGGGGAACGGCACTGGGTACCGTGGGTGGCGGACGACCTCAGCCTCGGGCCCTCAGCTGC 921  
QY 646 CTTTATCTGGACCTCCACCTCTTTCAGGAGAAATTAACAACATGTTATGAGAGGATGCT 705  
Db 922 CTGGCCTTGGAACTCCGATCTGCTTACAGAGCTGCACTGGACTCCGTGGCGCGCGCG 981  
QY 706 GAAACCCCATGGGATGGGGAAACAAATTTCTGCAAGAACCCAGATGGGACGAAAGCC 765  
Db 982 GCCCTGTGGGCTGGGCCCCCATGCTTACTGCGGAATCCGGACAATGACGAGAGGCC 1041  
QY 766 TGGTGCTTTTAAAGTTACCAATGACAAGGTGAATGGGAATCTATGATGCTCAGCC 825  
Db 1042 TGGTGCTACGT---GGTGAAGGACAGCGCGCTCTCTCTGGGAGTACTGCCCGCTGGAGGCC 1098  
QY 826 TGCTCAGCCAGGACGTGCTCTACCCAGGAGAAAGCCCACTGAGCCATCAACCAAGCTT 885

Db 1099 TCGAATCCCT-----CACCAGAGTCCAACTGTCAACCGATCTCTCGCGACCCCTG 1149  
QY 886 CCGGGTTTGAATCTCTGTGAAAGACTGAGATAGCAGAGAGAGATCAAGAGATCTAT 945  
Db 1150 CTTGAGCCAGCCTCCCGGGCGCCGAGCCTCTCGCAGGAGGACCAAGAGAGAGCTTC 1209  
QY 946 GGAGGCTTTAAGACGACGCGGGCAAGCAACCATCGCAGGGCTCCCTCCAGTCTCGTG 1005  
Db 1210 CTGGCGCCAGGTATCATCGGCGCTCTCTCTCGCTGCGCGCTCGCACCCCTG----- 1262  
QY 1006 CTTCTGACCATCTCCATGCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACCCCTGC 1065  
Db 1263 --GCTGGCGCCATCTACATCGGGGACAGCTTCTCGCGGGAGCCTGTGTCTCACACCTGC 1320  
QY 1066 TGGTGCTCACTGCTGCCCACTGACCGACATAAAACCAGACATCTAA-----AGGTG 1119  
Db 1321 TGGTGCTGTGGCGCCGCCACTCTTCTCAACCGACGAGAGATGAGATTCGCCAATATATGCA 1380  
QY 1120 GTCTAGGGGACCCAGGACCTGAAGAAAGAAATTTATAGCAGAGCTTTAGGTGGAG 1179  
Db 1381 GTCTGGCGCAGCACTTCTTCAACCGACGAGAGCTGACGACGACCTTCGCGATCGAG 1440  
QY 1180 AAGATATTCAAGTACAGCCACTACATGAAGAGATGAGATTCGCCAATATATGCA 1239  
Db 1441 AAGTACATCCGCTACACCTGTACTCGGTGTTCAACCCAGCGACCC---GACCTCGTC 1497  
QY 1240 TTGCTCAAGTTAAAGCCAGTGGATGTCTACTGTCTTCTAGAAATPCCAAATAGTGAAGCT 1299  
Db 1498 CTGATCGGCTGAAGAAAGAGGACCGCTGTGCCACACGCTCGCAGTTCTGTGACGCC 1557  
QY 1300 GTGTGCTGCTGAT-----GGGTCTTTCCTCTGGGAGTGGTGCACATCTCTGGC 1353  
Db 1558 ATCTGCTGCCGAGCCCGCAGCAGCCTTCCCGCAGGACACAAGTGCAGATTTGGGGC 1617  
QY 1354 TGGGG-----TGTTACAGAAACAGGAAAGGGTCCCGCAGCTCTCTGGATGCCAAAGTC 1407  
Db 1618 TGGGGCCACTTGGATGAGACGTGAGCGGCTACTCCAGCTCTCCGCGGAGGCGCTGGTC 1677  
QY 1408 AAGTGTATGCGCAACATTTGTGCAACTCCCGCAACTCTATGACACCATGATGATGAC 1467  
Db 1678 CCCCTGGTCCGCGACCAACAGTGCAGCAGCCCTGAGGTCTACGGCGCCGACATCAGCCCC 1737  
QY 1468 AGTATGATCTGTGAGGAATCTTCAGAAACCTGGGCAAGACACCTGCCAGGGTGAATCT 1527  
Db 1738 AACATGCTGTGTGCGGCTACTT---CGACTGCAAGTCCGACGCTGCCAGGGGACTCA 1794  
QY 1528 GGAGGCGCCCTGACCTGTGAGAGGACGCGACCTTACTGCTCTATGGGATAGTGGTGG 1587  
Db 1795 GGGGGCGCCCTGGCTTCGAGAGAGAGCGGCTGGCTTTACCTCTACGGCATCATCAGCTGG 1854  
QY 1588 GGCCTGGAGTGTGG-----AAGAGCGCAGGGGTCTACACCCCAAGTTACCAAAATTCCTG 1641  
Db 1855 GGTGACGGCTCGGGCGGCTCCCAAGCGGGGTCTACACCCGCGTGGCCACTATGTG 1914  
QY 1642 AATTGGATCAAGGCCACCAT 1661  
Db 1915 GACTGGATCAAGACCGGAT 1934

Search completed: May 25, 2004, 06:30:47  
Job time : 479 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 05:42:15 ; Search time 3059 Seconds  
(without alignments)  
16429.568 Million cell updates/sec

Title: US-09-912-559-1  
Perfect score: 1683  
Sequence: 1 atgtgttcagagatctctga.....aaagtgaagtggtttctaa 1683

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :	EST:*
1:	em_estba.*
2:	em_esthum.*
3:	em_estm1.*
4:	em_estm2.*
5:	em_estov.*
6:	em_estpl.*
7:	em_estro.*
8:	em_htc.*
9:	gb_est1.*
10:	gb_est2.*
11:	gb_htc.*
12:	gb_est3.*
13:	gb_est4.*
14:	gb_est5.*
15:	em_estfun.*
16:	em_estom.*
17:	em_gss_hum.*
18:	em_gss_inv.*
19:	em_gss_pln.*
20:	em_gss_vrt.*
21:	em_gss_fun.*
22:	em_gss_mam.*
23:	em_gss_mus.*
24:	em_gss_pro.*
25:	em_gss_rod.*
26:	em_gss_phg.*
27:	em_gss_vrl.*
28:	gb_gss1.*
29:	gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848.4	50.4	1042	13	BX325239
2	767.4	45.6	1008	13	BX463023
3	698.2	41.5	891	12	BI761782
4	647	38.4	889	13	BX431866

C	5	572.8	34.0	576	14	CB156834	K-EST0215
	6	537	31.9	663	14	CB162144	K-EST0222
	7	455.8	27.1	736	14	CB594245	AGENCOURT
	8	451.6	26.8	592	10	BE032018	130809
	9	447.2	26.6	802	12	BG972579	602841165
	10	437.8	26.0	674	9	AA217892	mw54508.r
	11	435.4	25.9	751	12	BI148082	602912432
	12	435.2	25.9	570	12	BM508620	ii37h03.Y
	13	416.4	24.7	929	10	BF788188	602113411
	14	399.4	23.7	665	14	CF169114	B0809C08-
	15	396.4	23.6	791	12	BI332440	60298081-
	16	395.8	23.5	546	9	AA237499	mx10810.r
	17	392.6	23.3	608	9	AV601564	AV601564
	18	390.4	23.2	902	10	BF780971	602105493
	19	386	22.9	682	12	BI220028	602934913
	20	384.6	22.9	529	10	BF785781	602112402
	21	379.8	22.6	648	14	CF171046	E0837E02-
	22	372.4	22.1	969	10	BF384535	602046804
	23	370.6	22.0	654	10	AW475402	un65h07.Y
	24	370.4	22.0	587	12	BG972681	602839039
	25	365.6	21.7	659	10	AM610902	un13g10.Y
	26	363.6	21.6	594	12	BM503097	ih51d12.Y
	27	363.2	21.6	625	14	CF171125	E0838E07-
	28	348.8	20.7	716	10	BF789705	602103735
	29	335.8	20.0	655	10	BM569555	BB569555
	30	335.4	19.9	668	13	BY742633	BY742633
	31	332.6	19.8	733	14	CB599445	AGENCOURT
	32	330.8	19.7	934	12	BI765113	603051314
	33	327.8	19.5	345	14	T68666	YC43e09.r1
	34	323	19.2	787	14	CB955374	AGENCOURT
	35	319	19.0	486	9	AA268125	vb08g12.r
	36	311.8	18.5	600	10	BF788541	602114406
	37	304	18.1	335	14	R89458	YQ01a10.r1
	38	293.2	17.4	304	14	T77362	Yd72g02.r1
	39	286.2	17.0	610	14	CA944054	in05e08.Y
	40	279.8	16.6	521	12	BI221826	602936942
	41	275.8	16.4	912	12	BG972626	602836922
	42	275.6	16.4	366	14	T93666	Ye06a11.r1
	43	266	15.8	434	14	T84369	Yd37g04.r1
	44	258.6	15.4	781	10	BF533788	602075308
	45	254.6	15.1	458	14	R10295	Yf36e12.r1

ALIGNMENTS

RESULT 1  
BX325239 1042 bp mRNA linear EST 01-MAY-2003  
LOCUS BX325239 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED  
DEFINITION Homo sapiens cDNA clone CS00J010124 5-PRIME, mRNA sequence.  
ACCESSION BX325239  
VERSION BX325239.1 GI:30309195  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1042)  
AUTHORS Li, W.B., Gruber, C., Jesses, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5634.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0AJ010DF12QP1  
&cluster=5634.f. Contact : Feng Liang Email : fliang@lifetech.com  
URL : http://fulllength.invitrogen.com/InvitroGen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0AJ010DF12QP1.

Db	925	AGTTACCAATGACAAAGTGAATGTGTGATGTCTGACGCTGCTCAGCCAGGA	984
QY	840	COTTGCTTACCAGAGAAAGCCCACTGAGCATCAACCAAGCTTCGGGGTTGACTC	899
Db	985	COTTGCTTACCAGAGAAAGCCC--ACTGAGCATCAACCAAGCTTCGGGGTTTACTC	1042
RESULT 2			
LOCUS	BX463023	1008 bp	linear
DEFINITION	BX463023 Homo sapiens FETAL LIVER Homo sapiens cDNA clone		
ACCESSION	CS0DM004YK02 5-PRIME, mRNA sequence.		
VERSION	BX463023.1	GI:31025478	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage Bp 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5634.f For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CS0DM004BF01Q1&cluster=5634.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DM004BF01Q1.		
FEATURES			
	Location/Qualifiers		source
	1..1008		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CS0DM004YK02"		
	/tissue type="FETAL LIVER"		
	/dev stage="fetal"		
	/clone_lib="Homo sapiens FETAL LIVER"		
	/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."		
ORIGIN			
	Query Match	45.6%;	Score 767.4; DB 13; Length 1008;
	Best Local Similarity	93.4%;	Pred. No. 1.2e-207;
	Matches	840;	Conservative 31; Mismatches 19; Indels 9; Gaps 7;
QY	1	ATGTTTGCAGGATGCTGATCTCCATGTTCTGCTGTTAAATGGCTTGTGGTGGAAAGACA	60
Db	110	ATGTTTGCAGGATGCTGATCTCCATGTTCTGCTGTTAAATGGCTTGTGGTGGAAAGACA	169
QY	61	GCTGTGGGTTCTCCCTGATGCTTTTATTTGAAAGCTTGGACCCAGACTGACCCCTG-A	119
Db	170	GCTGTGGGTTCTCCCTGATGCTTTTATTTGAAAGCTTGGACCCAGACTGACCCCTGAA	229
QY	120	CCAGTATGATTACAGTACGAGGATTAATATCAGGAGAGAACACCACTAGCACACTTAC	179
Db	230	CCAGTATGATTACAGTACGAGGATTAATATCAGGAGAGAACACCACTAGCACACTTAC	289
QY	180	CCATGCT-GAGAAATCCTGAGTGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCA	238
Db	290	CCACGCTAGAGAATCCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCA	349
QY	239	ACCCCTGTGTAACACCGTGGGACTGCTCGTCCATGGGAGCACTTCCATGCTGCTGCC	298



```
Db 350 ACCCTGTGAACACGGTGGGAGTGCCTGCTCATGGGAAACCTTTCACATGACGTGCC 409
Qy 299 TGGCTCTTTCTCTGGGAATAAGTG--TCAGAAAGTGCAGAAATACGTGCAAGCAACCC 356
Db 410 TGGCTCTTTCTCTGGGAATAAGTGATCABAHARATGCAGAAATACGTGCAAGCAACCC 469
Qy 357 ATGTGGCGGGGCAATGT--CTCATACCCAGAGTCTCCCTACTACCGTGTGTCTGT 414
Db 470 ATGTGGCGGGGCAATGTATCATATACCCAGAGTCTCCCTACTACCGTGTGTCTGT 529
Qy 415 AACACACCTTTACACAGGTCCAG--CTGCTCCCAAGTGGTTCCTGTATGACGCCCAACCC 473
Db 530 HAACACCTTTACACAGGTCCAGACTGCTCCCAAGTGGTTCCTGTATGACGCCCAACCC 589
Qy 474 CTGCCAGATGGGGTACTCTCTCCGGCATAGCCGAGATCCAGATTCACCTGTGCCTG 533
Db 590 CTGCCAGATGGGGTACTCTCTCCGGCATAGCCGAGATCCAGATTCACCTGTGCCTG 649
Qy 534 TCCGACACAGTCAAGGGGAAATCTGTGAAATAGGTTCTGTGACTGTATGTTGGCA 593
Db 650 TCMGACAGTCAAGGGGAAATCTGTGAAATAGTTCGTGACTGTATGTTGMA 709
Qy 594 TGGTACTCTTACCGAGGAAATGATAGGACAGTCAACCCAGCATGCGTCTTACTG 653
Db 710 TGGTACTCTTACCGAGGAAATGATAGGACAGTCAACCCAGCATGCGTCTTACTG 769
Qy 654 GAATCTCCACTCTCTTTCAGGAGATTAACACATGTTTATGGAGGATGCTGAAACCA 713
Db 770 GAATCTCCACTCTCTTTCAGGAGATTAACACATGTTTATGGAGGATGCTGAAACCA 829
Qy 714 TGGATTTGGGAAACACAAATTTCTGCAGAACCCAGATCGGAGCAAAAGCCCTGG--TGCT 772
Db 830 TGGATTTGGGAAACACAAATTTCTGCAGAACCCAGATCGGAGCAAAAGCCCTGGATGCT 889
Qy 773 TTATTAAGTTTACCAATGACAGGTGAATGGGATGATGATGATCTCAGCCTGCTCA 831
Db 890 TTATTAAGTTTACCAATGACAGGTGAATGGGATGATGATGATCTCAGCCTGCTCA 949
Qy 832 GCCCAGACGTTGCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGGG 890
Db 950 ACCCAGACGTTGCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGGG 1008
```

```
RESULT 3
BI761782 891 bp mRNA linear EST 25-SEP-2001
LOCUS 603046775F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5187066 5',
DEFINITION mRNA sequence.
ACCESSION BI761782
VERSION BI761782.1 GI:15753360
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 891)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: coapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M11467 Row: e Column: 19
High quality sequence stop: 839.
Location/Qualifiers
1..891
```

```
Query Match
Best Local Similarity 41.5%; Score 699.2; DB 12; Length 891;
Matches 775; Conservative 0; Mismatches 13; Indels 9; Gaps 6;
Qy 1 ATGTTTGCAGGATGTCGATCTCCATGTTCTGCTGTTAAATGGCTCTGGTGGAAAGACA 60
Db 81 ATGTTTGCAGGATGTCGATCTCCATGTTCTGCTGTTAAATGGCTCTGGTGGAAAGACA 140
Qy 61 GCCTGTGGGTTCTCCCTGATGTTTATTTGAAAGCTGGACCCAGACTGGACCCCTGCAC 120
Db 141 GCCTGTGGGTTCTCCCTGATGTTTATTTGAAAGCTGGACCCAGACTGGACCCCTGCAC 200
Qy 121 CAGTATGATTACAGCTACGAGGATTAATATCAGGAGAGAACACCACTAGACACTTACC 180
Db 201 CAGTATGATTACAGCTACGAGGATTAATATCAGGAGAGAACACCACTAGACACTTACC 260
Qy 181 CATGCTGAGAAATCCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCCAAC 240
Db 261 CAGCTGTGAAATCCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCCAAC 320
Qy 241 CCCTGTGAACAGGTGGGACTGCTCTGTCATGAGGAGCACTTCACATGACGTGCTG 300
Db 321 CCCTGTGAACAGGTGGGACTGCTCTGTCATGAGGAGCACTTCACATGACGTGCTG 380
Qy 301 GCTCTCTTCTCTGGGAATAAGTGTGAGAAATGCAAAATACGTGCAAGGACCAACCCATGT 360
Db 381 GCTCTCTTCTCTGGGAATAAGTGTGAGAAATGCAAAATACGTGCAAGGACCAACCCATGT 440
Qy 361 GGCGGGGGCAATGTCCTCATTTACCCAGAGTCTCTCCCTACTACCGCTGTGTGTTAAACAC 420
Db 441 GGCGGGGGCAATGTCCTCATTTACCCAGAGTCTCTCCCTACTACCGCTGTGTGTTAAACAC 500
Qy 421 CTTTACACAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGCCAAACCCCTGCCAG 480
Db 501 CTTTACACAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGCCAAACCCCTGCCAG 560
Qy 481 AATGGGGTACTGCTTCCCGGCAATGAGCGAGATCCAGATTCACTGTGCTGCTGCCGAC 540
Db 561 AATGGGGTACTGCTTCCCGGCAATGAGCGAGATCCAGATTCACTGTGCTGCTGCCGAC 620
Qy 541 CAGTTCAGGGGAAATTTCTGTAATAGTTCCTGATGACTGCTATGTTGGCGATGGCTAC 600
Db 621 CAGTTCAGGGGAAATTTCTGTAATAGTTCCTGATGACTGCTATGTTGGCGATGGCTAC 680
Qy 601 TCTTACCGAGGAAATGATAGGACAGTTC--AACAGATGCGTGCCTTTTACTGGAATC 659
Db 681 TCTTACCGAGGAAATGATAGGACAGTTC--AACAGATGCGTGCCTTTTACTGGAATC 740
Qy 660 CC--ACCTCTCTTTCAGGAGAAATTACACATGTTTATGGAGG----ATGCTGAACCCCATG 715
Db 741 CCAGCTCTCTTTCAGGAGAAATTACACATGTTTATGGAGGAGTCTTGAACCCCATG 800
Qy 716 GATTTGGGAA--CACAATTTCTGAG--AAACCCAGATGCGGAG--AAAGCCCTGTGTGC 771
```

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5187066"
/lab_host="DH108"
/clone_lib="NIH_MGC_116"
/notes="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."
```

ORIGIN

FEATURES
source

```

Db      801 GGATTGGGAAACACCAATTTCTGCAGAAAACCCAGATGCCGACGAAAAGCCCTGGTGC 860
QY      772 TTTATTAAAGTTACCAA 788
Db      861 TTTATTAAAGTCAACCA 877

RESULT 4
BX431866
LOCUS   BX431866 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
DEFINITION
CS0DM004YK02 5-PRIME, mRNA sequence.
ACCESSION BX431866
VERSION   BX431866
KEYWORDS EST.
SOURCE   BX431866.1 GI:30789052
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5634.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAG065ZH05_CS06184_1&cluster=5634.f.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAG065ZH05_CS06184_1.
FEATURES             Location/Qualifiers
     source            1..889
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="CS0DM004YK02"
                        /tissue_type="FETAL LIVER"
                        /dev_stage="fetal"
                        /clone_lib="Homo sapiens FETAL LIVER"
                        /note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
                        was primed with a NotI-oligo(dT) primer. Five prime end
                        enriched, double-strand cDNA was digested with Not I and
                        cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                        vector. Library was not normalized."

ORIGIN
Query Match      38.4%; Score 647; DB 13; Length 889;
Best Local Similarity 91.3%; Pred. No. 3e-173;
Matches 697; Conservative 0; Mismatches 65; Indels 1; Gaps 1;

QY      696 GGAGATGCTGAAGAACCCATGGGATGGGGAACACAAATTTGCGAGAAACCCAGATGCGGA 755
Db      1 GGAGATGCTGAAGAACCCATGGGATGGGGAACACAAATTTGCGAGAAACCCAGATGCGGA 60
QY      756 CGAAAAGCCCTGGTCTTTAATTAAGTTACCAATGACAAGTGGAATGGGAATCTGTGA 815
Db      61 CGAAAAGCCCTGGTCTTTAATTAAGTTACCAATGACAAGTGGAATGGGAATCTGTGA 120
QY      816 TGTCTCAGCCTGCTCAGCCAGGACGTTGCTTACCCAGAGAAAAGCCCCACTGAGCCATC 875
Db      121 TGTCTCAGCCTGCTCAGCCAGGACGTTGCTTACCCAGAGAAAAGCCCCACTGAGCCATC 180
QY      876 AACCAAGCTTCGGGGTTTGACTCTCTGTGGAAGAACTGAGATACAGAGGAGATCAA 935
Db      181 AACCAAGCTTCGGGGTTTGACTCTCTGTGGAAGAACTGAGATACAGAGGAGATCAA 240
QY      936 GAGAACTATGAGGCTTTAAGAGCACGCGCGGCAAGCACCCATGGCAGGCGTCCCTCCA 995

```

```

Db      241 GAGAATCTATGAGGCTTTAAGAGCAGCGGGGCAAGACCCCATGAGCGGCTCCCTCCA 300
QY      996 GTCTCGCTGCTCTGACCATCTCCATGCGCCAGGCGCCACTTCTGTGGTGGGCGCTGAT 1055
Db      301 GTCTCGCTGCTCTGACCATCTCCATGCGCCAGGCGCCACTTCTGTGGTGGGCGCTGAT 360
QY      1056 CCACCCCTGCTGGGTGCTCACTGCTGCCCACTCCACGACATPAACCAACAGACATCTAA 1115
Db      361 CCACCCCTGCTGGGTGCTCACTGCTGCCCACTCCACGACATPAACCAACAGACATCTAA 420
QY      1116 GGTGGTCTTAGGGGACAGGACCTGAAGAAAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGT 1175
Db      421 GGTGGTCTTAGGGGACAGGACCTGAAGAAAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGT 480
QY      1176 GGAGAAGATATTCAAGTTCAGTCACTACATGAAGAGATGAGATTTCCCAATGATAT 1235
Db      481 GGAGAAGATATTCAAGTTCAGTCACTACATGAAGAGATGAGATTTCCCAATGATAT 540
QY      1236 TGCATTGCTCAAGTTAAAGCCAGTGGATGCTCACTGTCTCTAGAAATCCAAATACGTGAA 1295
Db      541 TGCATTGCTCAAGTTAAAGCCAGTGGATGCTCACTGTCTCTAGAAATCCAAATACGTGAA 600
QY      1296 GACTGTGCTGTGCTCATGGGTCCTTCCCTCTGGAGTGAGTGCCACAT-CTCTGGCT 1354
Db      601 GACTGTGCTGTGCTCATGGGTCCTTCCCTCTGTAGAGAGTGCCACATGTTGTGACT 660
QY      1355 GGGGTGTACAGAAACAGGAAAGGGTCCCGCCAGCTCTCTGGATGCCAAAGTCAAGTGA 1414
Db      661 GTGGGTGTGAGAAACAGTTTAAGTTTCTTCCGCTACTGTTCCCTTAGATAGTTTCG 720
QY      1415 TTGCAACACATTTTGTGCAACTCCCGCAACTCTATGACCACAT 1457
Db      721 TTGTCGACATGGAGGGGGGGGGTGGTGGTACTGTGTGTGCACAT 763

RESULT 5
BX156834/c
LOCUS   BX156834 Homo sapiens cDNA clone
DEFINITION
K-EST0215807 L17N670205n1 Homo sapiens cDNA sequence.
ACCESSION BX156834
VERSION   BX156834
KEYWORDS EST.
SOURCE   BX156834.1 GI:28141962
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Choi, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
TITLE   21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 4 row: F column: 07
High quality sequence stop: 576.
Location/Qualifiers
     source            1..576
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="L17N670205n1-4-F07"
                        /sex="F"
                        /lab_host="Top10F"
                        /clone_lib="L17N670205n1"
                        /note="Organ: Liver; Vector: p7T3-Pac; Site 1: EcoRI;
                        Site 2: NotI; The library was contributed by the Soares

```

## ORIGIN

Query Match	34.0%	Score 572.8	DB 14	Length 576	
Best Local Similarity	99.7%	Pred. No. 3.8e-152			
Matches 574	Conservative	0	Mismatches 2	Indels 0	Gaps 0
494	QY	GCTCCCGCAGTAAGCGGAGATCAAGTTCACCTGTGCGCTGTCCCGACGAGTCAAGGGGA	553		
576	DB	GCTCCCGCAGTAAGCGGAGATCAAGTTCACCTGTGCGCTGTCCCGACGAGTTCAGGGGA	517		
554	QY	AATTCGTGTGAATAGGTTCTGATGACTGTCTATGCTGGCGATGGCTACTCTTACCGAGGGA	613		
516	DB	AATTCGTGTGAATAGGTTCTGATGACTGTCTATGCTGGCGATGGCTACTCTTACCGAGGGA	457		
614	QY	AAATGAATAGGACAGTCAACACGAGTCGCGTCTTTACTTGGAACTCCACACCTCTCTTGC	673		
456	DB	AAATGAATAGGACAGTCAACACGAGTCGCGTCTTTACTTGGAACTCCACACCTCTCTTGC	397		
674	QY	AGGAGAAATTCAAACATGTTTATGGAGGATGCTCAAACCCATGGGATTTGGGGAACAACAATT	733		
336	DB	AGGAGAAATTCAAACATGTTTATGGAGGATGCTGAACCCATGGGATTTGGGGAACAACAATT	337		
734	QY	TCTCAGAAAACCCAGATGCGGACGAAAAGCCCTGTGTCTTTATTAAGTTACCAATGACA	793		
336	DB	TCTCAGAAAACCCAGATGCGGACGAAAAGCCCTGTGTCTTTATTAAGTTACCAATGACA	277		
794	QY	AGGTGAATGGGAATCTGTGATGCTCAGCCCTGCTCAGCCAGGACGTTGCTTACCCAG	853		
276	DB	AGGTGAATGGGAATCTGTGATGCTCAGCCCTGCTCAGCCAGGACGTTGCTTACCCAG	217		
854	QY	AGGAAGCCCACTGAGGCCATCAACCAAGCTTCGGGGTTTGACTCTCTGTGAAGAGCTG	913		
216	DB	AGGAAGCCCACTGAGGCCATCAACCAAGCTTCGGGGTTTGACTCTCTGTGAAGAGCTG	157		
914	QY	AGATAGCAGAGAGGAAGATCAAGAGAATCTATGGAGGCTTTAAGACACGGCGGCAAGC	973		
156	DB	AGATAGCAGAGAGGAAGATCAAGAGAATCTATGGAGGCTTTAAGACACGGCGGCAAGC	97		
974	QY	ACCCATGCGAGCGTCCCTCCAGTCTCGCTGCTGACCTCTGCATCTCCATGCCCCCAGGGCC	1033		
96	DB	ACCCATGCGAGCGTCCCTCCAGTCTCGCTGCTGACCTCTGCATCTCCATGCCCCCAGGGCC	37		
1034	QY	ACTTCTGTGTGGGCGCTGATCCACCCCTGCTGGG	1069		
36	DB	ACTTCTGTGTGGGCGCTGATCCACCCCTGCTGGG	1		

RESULT 6	CB162144	663 bp	mRNA	linear	EST 30-JAN-2003
LOCUS	CB162144				
DEFINITION	K-EST02322650 LI17N670205n1 Homo sapiens cDNA clone				
	LI17N670205n1-17-H10 5', mRNA sequence.				
ACCESSION	CB162144				
VERSION	CB162144.1	GI:28148270			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1. (bases 1 to 663);				
AUTHORS	Kim,N.S., Hann,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.B., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.				
TITLE	21C Frontier Korean EST Project 2001				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Eoeun-dong Yuseong-qu, Daejeon 305-333, South Korea				

Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 17 row: H column: 10  
High quality sequence stop: 663.

## FEATURES

```

1.663
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L17N670205n1-17-H10"
/sex="F"
/lab_host="Top10F."
/clone_lib="L17N670205n1"
/note=Organ: Liver; Vector:
Site2: NotI; The library was
laboratory and it was constr
M.F., Lennon, G. and Soares,
6(9): 791-806. RNA was prepa
culture."

```

## ORIGIN

Query Match	31.9%;	Score 537;	DB 14;	Length 563;
Best Local Similarity	100.0%;	Pred. No. 7.3e-142;		
Matches 537;	Conservative	0;	Mismatches 0;	Indels 0; Gaps 0;
QY	1147	GAAGAATTTCATGACGACAGCTTTTAGGGTGGGAAGATATTCAACTGACAGCCACTACAAAT	1206	
DB	1	GAAGAATTTCATGACGACAGCTTTTAGGGTGGGAAGATATTCAACTGACAGCCACTACAAAT	60	
QY	1207	GAAGAGAGATGAGATTCGCCACAAATGATATTCGATTGCTCAAGTTAAAGCCAGTGGATGGT	1266	
DB	61	GAAGAGAGATGAGATTCGCCACAAATGATATTCGATTGCTCAAGTTAAAGCCAGTGGATGGT	120	
QY	1267	CACGTGTCTTAGAATCCMAATACGTGAAGACTGTGTGTCCTCGCTGATGGGTCCTTTCCC	1326	
DB	121	CACGTGTCTTAGAATCCMAATACGTGAAGACTGTGTGTCCTCGCTGATGGGTCCTTTCCC	180	
QY	1327	TCCTGGGAGAGTGGCCACATCTCGCTGGGGTGTTCAGAAACAGGAAAAAGGGTCCCCGC	1386	
DB	181	TCCTGGGAGAGTGGCCACATCTCGCTGGGGTGTTCAGAAACAGGAAAAAGGGTCCCCGC	240	
QY	1387	CAGCTCTCGGATGCCAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCGCCCAACTC	1446	
DB	241	CAGCTCTCGGATGCCAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCGCCCAACTC	300	
QY	1447	TATGACCAATGATTGATGACAGTATGATCTGTGCAGGAAATCTTCAGAAACCTGGGCAA	1506	
DB	301	TATGACCAATGATTGATGACAGTATGATCTGTGCAGGAAATCTTCAGAAACCTGGGCAA	360	
QY	1507	GACACTGCCAGGGTGACCTCTGGAGCCCTCTGACCTGTGAGAAGACGGCACTACTACTAC	1566	
DB	361	GACACTGCCAGGGTGACCTCTGGAGCCCTCTGACCTGTGAGAAGACGGCACTACTACTAC	420	
QY	1567	GTCTATGGGATAGTGAGCTGGGGCTGTGAGTGTGGGAAGAGCCAGGGGTCTACACCCAA	1626	
DB	421	GTCTATGGGATAGTGAGCTGGGGCTGTGAGTGTGGGAAGAGCCAGGGGTCTACACCCAA	480	
QY	1627	GTTCACAAATTCCTGAATTTGGATCAAAAGCCACCAATCAAAAGTGAAGAGTGGCTCTTAA	1683	
DB	481	GTTCACAAATTCCTGAATTTGGATCAAAAGCCACCAATCAAAAGTGAAGAGTGGCTCTTAA	537	

RESULT 7	ACCESSION	ORGANIS
CB594245	VERSION	
LOCUS	KEYWORDS	
DEFINITION	SOURCE	

CB594245 736 bp mRNA linear EST 03-APR-2003  
AGENCOURT\_12929103 NIH\_MGC\_177 Mus musculus cDNA clone  
IMAGE\_30310462 5' mRNA sequence.

EST.  
Mus musculus (house mouse)  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; Mus.  
 1 (bases 1 to 736)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Dr. Michael Brownstein  
 cDNA Library Preparation: Michael Brownstein Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: NDCM92 row: d column: 23  
 High quality sequence stop: 558.  
 Location/Qualifiers  
 1. 736  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30310462"  
 /lab\_host="DH10B (Tl-phage-resistant)"  
 /clone\_lib="NIH\_MGC\_177"  
 /notes="Organ: liver; Vector: pDNR-LIB; Site.1: Sfil  
 (ggccattggcc); Site.2: Sfil (ggcgctctggcc); cDNA made  
 by oligo-dT priming and directionally cloned. 5' and 3'  
 adaptors were used in cloning as follows:  
 5'-AAGCAGTGGTATCACGAGATGGCCATTACGCCGG-3' and  
 5'-ATTTCAGGCGCAGGCGCCACATG-dT(30)NN-3'. Full-length  
 enriched library was constructed using the Clontech  
 Creator SMART kit and size-selected to contain the 0.5 kb  
 size fraction. Library created in the laboratory of M.  
 Brownstein (NIH, NIH). Note: this is a NIH\_MGC Library."

ORIGIN  
 Query Match 27.1%; Score 455.8; DB 14; Length 736;  
 Best Local Similarity 80.9%; Pred. No. 1.3e-118;  
 Matches 543; Conservative 0; Mismatches 127; Indels 1; Gaps 1;  
 715 GGGATTGGGNAACATATTTTCGCAAAACCAGATCGGACGAAAGCCCTGGTGT 774  
 Db 3 GGGATCGCAGAGCAACAATTTCTCGAAACCAGATGGAGACCAAAACCTGGTGT 62  
 QY 775 ATTAAGATTCAATGACAAAGGTGAATGGGAATCTGTGATGTCTAGCCTGCTCAGCC 834  
 Db 63 GTCAAGGTGAACAGTGAGAAGGTGAATGGGAATCTGTGATGTCTACAGTCTGCTCAGTG 122  
 QY 835 CAGACAGTTGCTTACCAGAGGAAACCCCACTGAGCCATCAACCAAGCTTCGGGGGTTT 894  
 Db 123 CTTGACACCCCTAAACCAAGTGGAAAGCCTTCTGGAGCCTGTGATGGAGCTGCCAGGGTTC 182  
 QY 895 GACTCCTGTGAAAGACTGAGATAGCAGAGAGGAAGATCAAGAGAAATCTATGGAGGTTT 954  
 Db 183 GAGTCTTCGGGAAGACGGAGGTAGCTGAACACGAGTCAAGCGTATCTACGGGGGTTT 242  
 QY 955 AAGAGCAGCGGGCAGCACCACCTGGCAGCGTCTCTCAGTCTCGCTGCCCTCAGCC 1014  
 Db 243 AAGAGCAGCAGGCAAGCACCCTGGCAGTGTCTTCAGACTCTACTCGCTTGACC 302  
 QY 1015 ATCTCCATGCCCCCAGGGCCCACTTCTGTGGTGGGGCGCTGATCACCCCTGCTGGGTGTC 1074  
 Db 303 ACCTCATGCCCCCAGGSCCACTTCTGTGGGGGGCGCCTGATCACCCCTGCTGGGTGTC 362  
 QY 1075 ACTGTGTCCTCCTGACCCGACATATAAAACACAGACATCTAAAGTGGTCTAGGGGACCAG 1134  
 Db 363 ACTGAGCCCACTGTACCGACATTAACACCAAGCATCTAAAGTTGTACTAGGGATCAG 422  
 QY 1135 GACCTGAAAGAAGAAGAAATTCATGACGACAGCTTTAGGGTGGAGAGATATCAAGTAC 1194  
 Db 423 GACCTGAAAGAAGACAGAATCCCATGAACAGACCTTCAGGGTGGAAAAAATACTGAAGTAC 482

```
Db 66 CTATGTTGGTACCGGCTACTCTTACCGAGGGAAGTGAATAACTGTCAACGACAC 125
Qy 642 GTGCTTTTACTGGAACCTCCACCTCTCTTTCAGAGGAATTTACAACTGTTATGGAGGA 701
Db 126 GTGCTTTTACTGGAACCTCCACCTCTCTTTCAGAGGAATTTACAACTGTTATGGAGGA 185
Qy 702 TGCTGAACCCATGGATTTGGGAACAAATTTCTGCAGAAACCCAGATCGGAGGAAA 761
Db 186 TGCTGAGCCCATGGATTTGGGAGCAAACTCTGCAGAAACCCAGATGGAGAGAAA 245
Qy 762 GCCCTGTGTCTTTTAAAGTTACCAATGACAAGGTGAAATGGAATGGAATGCTGTC 821
Db 246 GCCCTGTGTCTTTTAAAGTGAGCAGTGCCAAAGTGAAATGGAGTACTGTGACGTCC 305
Qy 822 AGCTGCTCAGCCAGGAGTGTCTACCCAGAGGAAGCCCACTGAGCCATCAACAA 881
Db 306 TGCTGTCTCAGC-----CCGAGAGGGAAGCCCACTAGAACCTGTGACCGA 350
Qy 882 GCTTCCGGGGTTTGACTCTGTGGAAGACTGAGATAGCAGAGGAAGATCAAGAAAT 941
Db 351 GCTTCCCGGGTTTGACTCTGTGGGAGCAGACAGCAGAAAGGAAGTTCAAGAGAT 410
Qy 942 CTATGGAGGCTTTAAGACAGCGGGGCAAGCACCATGCGAGGGTCCCTCCAGTCTC 1001
Db 411 CTACGAGGCTTTAAGACAGCGGGGCAAGCACCCTGGCAGCGCTCCCTACAGACCTC 470
Qy 1002 GCTGCTCTGACCACTCTCATGCCCGGCGCACTTCTGTGGTGGGGGCTGTATCCACC 1061
Db 471 CTGATCTGACCGTCTCATGCCCGGCGCACTTCTGTGGGCGGCTGTATCCACC 530
Qy 1062 CTGCTGGGTGCTCACTGTGCCACTGCAACCGACATAAAACCAATCAAGGTGGT 1121
Db 531 CTGCTGGGTGCTCACTGTGCCACTGCAACCGACATAAAAGCCAAATATCTAAAGTAGT 590
Qy 1122 GC 1123
Db 591 GC 592

RESULT 9
Bg972579 802 bp mRNA linear EST 12-JUN-2001
LOCUS 602841165F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4975394
DEFINITION 5', mRNA sequence.
ACCESSION Bg972579
VERSION Bg972579.1 GI:14360216
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M10967 row: a column: 17
High quality sequence stop: 785.
Location/Qualifiers
1..802
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
```

## FEATURES

source

RESULT 10  
AA217892  
LOCUS  
DEFINITION mv54b08.r1 Soares mouse 3NBE12 5 Mus musculus cDNA clone  
IMAGE:658839 5, similar to TR:G1345398 G1345398 HGF ACTIVATOR LIKE  
PROTEIN. ;, mRNA sequence.  
ACCESSION AA217892

/clone="IMAGE:4975384"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI\_CGAP\_Kid14"

Note="Origin: Kidney; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.75 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN

Query Match 26.6%; Score 447.2; DB 12; Length 802;  
Best Local Similarity 77.7%; Pred. No. 48-116;  
Matches 580; Conservative 0; Mismatches 158; Indels 8; Gaps 3;

Qy 1 ATGTTTGCAGGATGTCGATCTCCATGTTCTGCTGTTTAAATGGCTCTGGTGGAAAGACA 60  
Db 50 AATATTGTGAGATGTTGGTGTGCTGCTGCTGCTAATCGCTGCTGGTGGGAAGTCA 109  
Qy 61 GCTGTGGGTTCTCCCTGATGTTCTTATTTGAAAGCTGGACCCAGACTGGACCCCTGAC 120  
Db 110 GTCATTGGGCTCTCACTGATGTCCTTCAITGCGCCGCCAGACCCAGATTGGACCCCGAT 169  
Qy 121 CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAGAACACCACTAGTAGCACACTTACC 180  
Db 170 GACTATTACTACAGCTATGAGCAGTCCAGCCAGACGAGACCCCGAGTGTCAAGCAGACC 229  
Qy 181 CATGTGGAATCTGACTGTGTTACTACTGAGGACCAAGCTGATCCATGCGAGGCCAC 240  
Db 230 ACCCTTGAAACCCCGACTGCTACT-----ATGAGAGCATGATCCATGCCAGTCCAAC 283  
Qy 241 CCTGTGAACAGGTGGGACTGCTGCTCCATGGAGACCTTCACATGACAGCTGCCCTG 300  
Db 284 CCTGTGAACAGCGCGGAGCTGTATCATCAGAGGGGATACCTTCAGTTGAGCTGCCCA 343  
Qy 301 GTCCTCTTCTCTGGGAATTAAGTGTGAGAAAGTGCAAAATACGTGCAAGGACAAACCATGT 360  
Db 344 GCCCTCTTCTCGGGAGCGGTGCCAGACTGCCACAAAACAAGTGAAGGACAAACCATGT 403  
Qy 361 GCGGGGCGCAATGTCCTATTACCCAGAGTCTCCTACTACCGCTGTCTGTGTAACAC 420  
Db 404 GTCCATGTTGATTGCTCTATTACCCAGAGCACCCTTACTACCGCTGTCTGTGTAACAC 463  
Qy 421 CCTTACACAGTCCAGCTGCTCCCAAGTGTTCCTGTATGCAAGGCAAAACCCCTGCCAG 480  
Db 464 CCTTACACGGGACAGACTGCTCCAAAGTGTTCGGGATGCAAGGCAAAACCCCTGCCAG 523  
Qy 481 AATGGGGTACCTGCTCCCGGATAGCGGAGATCCAGTTCACTGCTGCTGCCGAC 540  
Db 524 AATGGCGGAGTCTGTTCCCGACACAGCAGGAGATCCAGTTTACCTGTGCTGCTCCAGAC 583  
Qy 541 CAGTTCAAGGGGAAATTTCTGTAA-ATAGGTTCTGATGACTGCTATGTTGGGATGGCTA 599  
Db 584 CAGTATAAGGGGAAATTTCTGTGAACATAGTCCGACGACTGTTATGTGCGTATGGCTA 643  
Qy 600 CTCTTACCGAGGGAATAATGATAGGACAGTCAACAGCAGATGCGTCTTACTGGAACTC 659  
Db 644 CTCTTACCGAGGCAAAAGTGAAGACAGTCAACACAGAACCCAGGSCCTTTACTGGAACT- 702  
Qy 660 CCACCTCTCTTCGAGGAGATTTACAACATGTTTATGGAGGATGCTGAAACCCATCGGAT 719  
Db 703 CCACTCTCTTCGAGGAGACTTTATACATGTTTATGGAGGATGACAGACCCACCGGAT 762  
Qy 720 TGGGGAACACAATTTCTGAGAAACC 745  
Db 763 GCGAGAGCACAACTTCTGCAAGAAC 788

```

VERSION AA217892.1 GI:1826975
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 674)
Marral,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,D., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilton,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:404687
Possible reversed clone: similarity on wrong strand
Seq primer: -28mi3 rev2 ET from Amersham
High quality sequence stop: 485.
FEATURES
Location/Qualifiers
1..674
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:658839"
/sex="unknown"
/tissue_type="fetus"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
/clone_lib="Soares mouse 3NHE12 5"
/notes="Organ: whole fetus; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Site 1: Not I; Site 2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5', TGTTACCAATCTGAAGTGGAGCGCGCCCTATTTTCTTTTCTTTT
3', on total mouse RNA (provided by Minoru Ko, Wayne
State Univ.); double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN
Query Match 26.0%; Score 437.8; DB 9; Length 674;
Best Local Similarity 78.2%; Pred. No. 1.8e-113;
Matches 526; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
QY 701 ATGCTGAACCCATGGGATGGGGAACACAAATTTCTGCAGAAACCCAGATGGGAGAAA 760
DB 1 ATTCGGAAGGCGACGGATCGCAGACACAACTTCTGCAGAAACCCAGATGGGAGACCA 60
QY 761 AGCCCTGCTGCTTTATTAAGTTACCAATGACAGGTGAATGGATCTGATGATGCT 820
DB 61 AACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
QY 821 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 880
DB 121 CAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 881 AGCTTCGGGGTTTATCTCTGCTGGAAGACTGATGATGATGATGATGATGATGATGATGAT 940
DB 181 AGCTGCCAGGGTTGAGTCTCTGCGGGAGAGCGGAGGTAGTGTGACACGCGAGTCAAGCGTA 240
QY 941 TCTATGAGGCTTTTAAAGAGCAGCGCGGAGACCCATGCCAGGCTCCCTCCAGTCCCT 1000

```

```

DB 241 TCTACGGGGGCTTTAAGAGCACAGCAGGCAAGCACCCTGGCAGGTGTCTCTGCAGACCT 300
QY 1001 CGCTGCTGTGACCATCTCCATGCCCCAGGGCCACTTCTGTGGTGGGGCGGTGATCCACC 1060
DB 301 CACTGCGGTGTGACCATCTCCATGCCCCAGGGCCACTTCTGTGGTGGGGCGGTGATCCACC 360
QY 1061 CTTGCTGGTGTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1120
DB 361 CTTGCTGGTGTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 1121 TGTCTAGGGGACAGGACCTGGAAGAAGAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGA 1180
DB 421 TACTAGGGGATCAGGACCTGGAAGAAGAGAGAGATCCCATGACAGACCTTCAGGGTGGAAA 480
QY 1181 AGATATTCAAGTACAGCCACTACATGAAGAAGATGAGATTTCCCAATGATATGATTCAT 1240
DB 481 AAATACTGAAGTACAGTCACTAGTATTAAGAAGAGATGAGATTTCCCAATGATATGATTCAT 540
QY 1241 TGTCTCAAGTAAAGCCAGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1300
DB 541 TGTCTCAAGTAAAGCCAGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
QY 1301 TGTCTTGTGCTGATGGGTCTTTTCCCTCTGGAGTGAAGTGCACATCTCTGCTGGGGTGG 1360
DB 601 TATGTTTGACAGCGACCCCTTTGCCCTCTGCACTGAAGTGCACATCTCTGCACTGGGTG 660
QY 1361 TTACAGAAACAGG 1373
DB 661 GTACAGAAACAGG 673

RESULT 11
BI148082 751 bp mRNA linear EST 05-JUL-2001
LOCUS 602912432F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5053663 5',
DEFINITION mRNA sequence.
ACCESSION BI148082
VERSION BI148082.1 GI:14608083
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 751)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M1145 row: 0 column: 08
High quality sequence stop: 747.
Location/Qualifiers
1..751
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5053663"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Li9"
/notes="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

```

Query Match	25.9%	Score 435.4	DB 12	Length 751
Best Local Similarity	78.0%	Pred. No. 9.2e-113		
Matches 589	Conservative 0	Mismatches 156	Indels 10	Gaps 5
QY	4	TTTGCAGGATGCTGATCTCCATGTTCTGCTTAATGGCTCTGCTGGAAAGACAGC- 62		
DB	2	TTTGTCAAGATGTTGGTTCCTGCTCCGTCTAATCGCCCTGGTGGGNACTCAGTG 61		
QY	63	CTGTGGGTTCTCCCTGATGTCTTTATTGGAAAGCTTGGACCCAGACTGGACCCCTGACCA 122		
DB	62	CATTGGGCTCTCACTGATGTCTTCATTGGCCCCCAGACCCAGATTGGACCCCGATGA 121		
QY	123	GTATTGATTACAGCTACGAGGANTTATAATCAGGAAGAGAACACAGTAGCACACTTACC 182		
DB	122	CTATTACTACAGCTATGAGCAGTCCAGGCCAGACGAAGACCCAGTGTACCGAGACCAC 181		
QY	183	TGCTGAGNATCTGACTGGTACTACTGAGGACCAAGCTGATTCATGCCAGCCCAACCC 242		
DB	182	CCCTGAGAACCCCACTGGTACT-----ATGAAGACGATGATTCATGCCAGTCCAAACC 235		
QY	243	CTGTGAACACGGTGGGACTGCCCTCCATCGGAGCACCTTCACATCAGCTGCTGCTGGC 302		
DB	236	CTGTGAACACGGGGGACTGTATCATCAGAGGGGATACCTTCAGTTCAGCTSCCCAGC 295		
QY	303	TCCTTTCTCTGGGAATTAAGTGTGAGAAAGTGGAATATGTCGPAAGGACCAACCATGTGG 362		
DB	296	CCCTTCTCTGGGGAGCCGGTGCAGACTGCAACAAACAAAGTCAAGGCAACCCATGTGTT 355		
QY	363	CCGGGGCCCAATGTCTCATTAACCCAGAGTCCCTCCTACTACCGCTGTGTCTGTAAACACC 422		
DB	356	CCATGGTGATTGCGCTCATTAACCCAGAGCACCCCTACTACCGCTGTGCTGCAATATACC 415		
QY	423	TTACACAGGTCACAGCTGCTCCCAAGTGTCTCTGTATGAGGCCCAACCCCTGCCAGAA 482		
DB	416	TTACACGGGACACAGCTGCTCCAAAGTCTTCGGCATGAGGCCAAACCCCTGCCAGAA 475		
QY	483	TGGGGCTACCTGCTCCCGGCATAAAGCGGAGATCCAAAGTTCACTGTGCTGTGCTCCCGACA 542		
DB	476	TGGCGGAGTCTGTTCCCGCACACAGACGAGATCCAGGTTTACTGTGCTGTGCTCCAGACA 535		
QY	543	GTTCAAGGGGNAATCTCTGAATAGGTCTGATGACTCTGATGTGTGGCATGGCTACTC 602		
DB	536	GTAAGAAGGGNAATCTCTGAATAGGTCTCGGACGACTGTTATGTCGGTGTATGGGCTACTC 595		
QY	603	TTACCGAGGGGAAAATGAATAGGACAGTCAACACGACATGCGTGCTCTTTACTGG- AACTCCC 661		
DB	596	TTACCGAGGCAAGTGAGTGAAGACAGTCAACACGAGACCCCATGCCCTTTACTGGAAACTCCC 655		
QY	662	ACCTCCTCTTGACGGGAATTAACAATGTTTATGAGGATGCTGAACCCCATGGGATTG 721		
DB	656	ACCTCCTCTTGACGAGACTTATAAACATGTTTATGAGGATGAGAGACCCACCGGA- TC 714		
QY	722	GGGAACCAATTTCTCGAGAAACCCAGATCGGAC 756		
DB	715	GAGAGCAACACTTCTG-AGAAAACCCAGATGGAGAC 748		

RESULT 12  
 BM508620  
 LOCUS  
 DEFINITION  
 570 bp mRNA linear EST 12-MAR-2002  
 .i137h03.v1 Melton Normalized Mixed Mouse Pancreas 1 N1-MMS1 Mus  
 musculus cDNA Clone IMAGE:5944156 S' similar to TR:Q14520 Q14520  
 HGF ACTIVATOR LIKE PROTEIN. ; mRNA sequence.  
 ACCESSION  
 BM508620  
 VERSION  
 BM508620.1 GI:18679763  
 EST.  
 Mus musculus (house mouse)  
 SOURCE  
 Mus musculus  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 570)  
 Melton, D., Brown, J., Keny, G., pernutt, A., Lee, C., Kaesner, K.,  
 Lenishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
 1

Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,  
 Schmitt, A., Thesling, B., Ritter, E., Ronko, I., Bennett, J.,  
 Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarisvilli, R.,  
 Williams, T., Jackson, Y., and Bowers, Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 Other ESTs: i137h03.x1  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@iobhp.harvard.edu  
 Library was constructed by Dr. Douglas Melton DNA sequencing by:  
 Washington University Genome Sequencing Center for information on  
 obtaining a clone please contact: Juliana Brown  
 (brownja@wustl.edu)  
 MGI:2007028 This sequence now available from the IMAGE consortium,  
 for clone orders contact: info@image.llnl.gov  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 435.

[illegible]

```

1. 570
Location/Qualifiers
/organism="Xus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:594156"
/sex="Both for embryonic & newborn, male for adult and
adult islet"
/dev stage="Embryonic day 10.5, E12.5, E16.5, newborn,
adult, mixed"
/lab_host="OH108"
/clone_lib="Melton Normalized Mixed Mouse Pancreas 1
NI-MMS1"
notes="Vector: pSPORT1; Site 1: Not I; Site 2: Sal I; Five
libraries representing E10.5/12.5 pancreatic bud, E16.5
pancreas, newborn pancreas, adult pancreas, and adult
islets of Langerhans were separately constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
was made by oligo-dT priming and size-selected by column
fractionation. Libraries were amplified once on solid
support and plasmid DNA from each library was prepared
and mixed in equal amounts. The mixed library DNA was
normalized by method #4 from Bonaldo, Lennon, and Soares
1996 Genome Research 6:791-806; 0.5 microgram
single-stranded mixed library plasmid DNA was mixed with
5 micrograms PCR product representing mixed library
inserts and hybridized to an EcoT 6. Single-stranded
(unhybridized) plasmids were isolated by hydroxyapatite
chromatography and used to make this library."

```

## ORIGIN

	Query Match	25.9%	Score 435.2	DB 12	Length 570
	Best Local Similarity	85.4%	Pred. No. 8.8e-113		
	Matches 485	Conservative 0	Mismatches 83	Indels 0	Gaps 0
QY	1108	CATCTAAAGTGTGCTAGGGGACCGAGGACCTGGAAGAAGAAGATTTTCATGACGACAGC	1167		
Db	3	CATCTAAAGTTGTACTAGGGGATCAGGACCTGGAAGAAGACAGAAATCCCATGAACAGACC	62		
QY	1168	TTTAGGGTGGAGAGATATTCAAGTACACCCACTACAATGAAGAGATGAGATTTCCCCAC	1227		
Db	63	TTCAGGGTGGAAATAATCTGAGTTACAGTCAGTATATATGAAGAGATGAGATTTCCCCAC	122		
QY	1288	AATGNAATTCGATTGCTCAAGTTAAAGCCAGTGGAGTGTCACTGTGCTCTAGAAATCCAAA	1287		
Db	123	AATGACATTCGTTTGTCTCAAGTTAAAGCCAGTGGGTGGTCACTGTGCTCTGGAATCCAGA	182		
QY	1288	TACGTGAAGACTGTGTGCTTTGCGCTGATGGTCTCTTTCCCTCTGGGAGTGTGCGCACATC	1347		



Db 183 TATGTGAAGACTGTATGTTGGCCAGCGACCCCTTTCCCTCTGGAACCTGAAGTGCACATC 242  
 Qy 1348 TCTGGCTGGGCTGTACAGAAACAGAAAGGTCGCCGACCTCTCTGGATGCCAAAGTC 1407  
 Db 243 TCTGGCTGGGCTGTACAGAAACAGGAGGTCGCCGACCTCTCTGGATGCCAAAGTC 302  
 Qy 1408 AAGCTGATTCGCAACACATTTGTGCAACTCCGCCAACTCTATGACCAATGATGATGAC 1467  
 Db 303 AAGCTAATCGTAAACCTTTGTGCAACTCCGCCAACTCTATGACCAACATCGATGAC 362  
 Qy 1468 AGTATGATCTGTGCAGGAATCTTCAGAAACCTGGCAAGACACCTGCCGGTGACTCT 1527  
 Db 363 AGTATGATTTGTGCGGGGAACCTTCAGAAAGCCCGGATCAGACACCTGCCAGGTGACTCG 422  
 Qy 1528 GGAGGCCCCCTGACCTGTGAGAGGACGCGACCTACTACTGCTATGAGGATAGTGAAGTGG 1587  
 Db 423 GGGGGCCCCCTAACTGTGAGAGGATGGAATTAACCTCTACCTCTACGGGATGTAAGCTGG 482  
 Qy 1588 GGCCTGGATGTGGGAAGGCGGAGGCTCTACCCAGTTTACCAATTTCTGATTTGG 1647  
 Db 483 GGCAGGAATGTGGGAAGGCGGAGGCTCTACCTCAAGTCACCAAGTTCTGATTTGG 542  
 Qy 1648 ATCAAGCCACCATCAAAAGTGAAGTG 1675  
 Db 543 ATTAAGACCACCATGCACAGGAGGCTG 570

RESULT 13  
 BF788188 929 bp mRNA linear EST 12-JAN-2001  
 LOCUS 602113411F1 NCI CGAP\_Kid14 Mus musculus cDNA clone IMAGE:4241642  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BF788188  
 VERSION BF788188.1 GI:12093224  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 929)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-rc@mail.nih.gov](mailto:cgapbs-rc@mail.nih.gov)  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LHAM9858 row: e column: 03  
 High quality sequence stop: 686.  
 FEATURES Location/Qualifiers  
 1..929  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4241642"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NCI\_CGAP\_Kid14"  
 /note="Organ: Kidney; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.75 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library. |"

FEATURES source  
 1..929  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4241642"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NCI\_CGAP\_Kid14"  
 /note="Organ: Kidney; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.75 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library. |"  
 ORIGIN  
 Query Match 24.7%; Score 416.4; DB 10; Length 929;  
 Best Local Similarity 75.9%; Pred. No. 2.9e-107;  
 Matches 529; Conservative 0; Mismatches 162; Indels 6; Gaps 1;  
 Qy 1 ATGTTGCCAGATGCTGATCTCATGTTCTGCTGTTAATGGCTCTGTTGGGAAGACA 60  
 Db 44 ATATTGTGAGATGATGATGTTCCGCTGCTGCTAATCGCTGTTGGGAAGTCA 103  
 Qy 61 GCCTGTGGGTCTCCCTGATGTCCTTATTGGAAAGCCTGGAGCCAGACTTGGACCCCTGAC 120  
 Db 104 GTATGGGCTCTCATGATGTCCTTATGCGCCCCAGAGCCAGATGGACCCCGCAT 163  
 Qy 121 CAGTATGATPACAGCTACGAGGATTAATCAGGAAGAGAACACAGTACGACACTTACC 180  
 Db 164 GACTATTACTACAGCTATGAGCAGTCCAGCCAGACGAAAGACCCAGTGTACGCGAGAC 223  
 Qy 181 CATGCTGAGAACTCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCGAGCCCAAC 240  
 Db 224 ACCCTGAGAAACCCGAGCTGTAATCT-----ATGAGAGCATGATCCATGCCAGTCCAAC 277  
 Qy 241 CCTGTGAACACACGGTGGGAGTCCCTGCTCATGAGGAGCACCTTCAATGACAGTCCCTG 300  
 Db 278 CCTGTGAACACACGGGAGTCTATCATCAGAGGGGATACCTTCACTTTCAGTTCAGCTGCCA 337  
 Qy 301 GCTCCTTCTCTGGGATAGTGTCAAGAGTCAAAATACGTGCAAGGACAAACCCATGT 360  
 Db 338 GCCCCTTCTCGGGAGCCGCTGCCAGATGACAAACCAAGTGCAGAGCAACCCATGT 397  
 Qy 361 GGCGGGGCCAATGTCTCATTTACCCAGAGTCTCCCTATACCGCTGTGTCTGTTAAACAC 420  
 Db 398 GTCCATGGTGTGCTCTCATTTACCCAGAGCAACCCCTACTACCGCTGTGCTGCAATAAC 457  
 Qy 421 CTTTACACAGTCCAGCTCTCCAGTGTCTTCTGTATGACAGGCCAAACCCCTGCGAC 480  
 Db 458 CTTTACACAGGACCCAGTCTCTCAAGTGTCTTCCGATGAGGACCAACCCCTGCCAG 517  
 Qy 481 AATGGGCTACCTGCTCCCGGCATTAAGCGAGATCCAAAGTTCACCTGTGCTGTCCCGAC 540  
 Db 518 AATGGCGAGTCTGTTCCCGACACAGACGAGATCCAGGTTTACCTGTGCTGTCCAGAC 577  
 Qy 541 CAGTTCAGGGGAAATCTGTGAATAGTCTTGTGATGACTGCTATGTTGGCGATGGCTAC 600  
 Db 578 CAGTATAAGGGGAAATCTGTGAATAGTCCGAGACACTGTTATGTCGGTGTATGGCTAC 637  
 Qy 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATCGCTGCTCTTACTGGAATCC 660  
 Db 638 TCTTACCGAGCAAGTGAATAGACAGTCCACAGAAACCATGCTTTACTGGACTCC 697  
 Qy 661 CACTCTCTTTCAGGAGGATTAACATGTTATGG 697  
 Db 698 CACTCTCTTTCAGGAGGACTTATCATGTTCTCTCG 734  
 RESULT 14  
 CF169114  
 LOCUS B0809C08-5 NIA Mouse Newborn Kidney cDNA Library (Long 1) Mus  
 DEFINITION musculus cDNA clone NIA:B0809C08 IMAGE:30468319 5', mRNA sequence.  
 ACCESSION CF169114  
 VERSION CF169114.1 GI:33278663  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 665)  
 AUTHORS Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.  
 TITLE Construction of long-transcript enriched cDNA libraries from  
 submicrogram amounts of total RNAs by a universal PCR amplification  
 method  
 JOURNAL Genome Res. 11 (9), 1553-1558 (2001)  
 MEDLINE 21429098  
 PUBMED 11544199  
 COMMENT Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA



Qy	181	CATGCTGAGAACTCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCCAAC	240
Db	203	ACCCCTTGAGAACCCCGACTGCTACT-----ATGAGACGATGATCCATGCCAGTCCAAAC	256
Qy	241	CCCTGTGAACACCGTGGGAGCTCCCTGCTCCATGGGAGCACCTTCACATGACAGTGCCTG	300
Db	257	CCCTGTGAACACCGCGGGGACTGTATCATCAGAGGGGATACCTTCAGTTGACAGCTGCCCA	316
Qy	301	GCTCCTTTCTCTGGGAATAAGTGTCAAGAAAGTGCAGAAATACGTGCAAGGACAAACCCATGT	360
Db	317	GCCCCCTTCTCGGGGAGCCGGTCCAGACTGCACAAAACAAGTGACAGGACAAACCCATGT	376
Qy	361	GGCCGGGGCCCAATGTCTCATTTACCCAGAGTCTCTCCCTACTACCGTGTCTCTGTAAACAC	420
Db	377	GTCCATGGTGATTCCTCTCAATTACCCAGAGGACACCCCTACTACCGTGTCTCTGTCAATAC	436
Qy	421	CCTTACACAGGTCCTCAGCTGCCCAAGTGGTTCTCTGTATGACAGGCCAAACCCCTGCCAG	480
Db	437	CCTTACACGGGACACAGCTGCTCCAAAGTCTTCCGGCATGACAGGCCAAACCCCTGCCAGA	496
Qy	481	AATGGGGCTACCTGCTCCCGGCATAAGCGGAGATCCAAAGTTCACTGTGCTCTGCCGAC	540
Db	497	ATGGCCGGAGTCTGTTCCTCCGACACAGACGAGATCCAGGTTTACCTGTGCTGTCCAGAC	556
Qy	541	CAGTTCAGGGGAAATCTGTGAA-ATAGTTCTGATGACTGCTATGTTGGGATGGCTA	599
Db	557	CAGTATAAGGGGAAATCTGTGAACATAGTCCGGACGACTGTTATGTCGGTGATGGCTA	616
Qy	600	CTCTTACCGAGGAAATGAATAGGACAGTCAACCCAGCATGCGTGCCTTTACTGGAACTC	659
Db	617	CTCTTACCGAGGCAAAAGTGAGTAAGACAGTTCAACAGAACCCATGCCTTTACTGGAACTC	676
Qy	660	CCAGCTCCTCTTGCAGGAGAAATTACAAC	687
Db	677	CCAGCTCCTCTTGCAGGAGACTTATAAC	704

Search completed: May 25, 2004, 10:42:51  
Job time : 3066 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 05:49:14 ; Search time 91.5 seconds  
(without alignments)  
10207.457 Million cell updates/sec

Title: US-09-912-559-1  
Perfect score: 1683  
Sequence: 1 atgtttgccaggatgtctga.....aaagtgaagtggtctcttaa 1683

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	117.6	7.0	2033	1	US-08-148-910-14
2	117.6	7.0	2033	1	US-08-448-937A-14
3	96.2	5.7	970	1	US-08-148-910-3
4	96.2	5.7	970	1	US-08-448-937A-3
5	75.4	4.5	1065	1	US-08-427-640-1
6	75.4	4.5	1065	1	US-08-427-640-5
7	75.4	4.5	1068	1	US-08-427-640-3
8	75.4	4.5	1137	4	US-09-553-498-9
9	75.4	4.5	1137	4	US-09-618-869-9
10	75.4	4.5	1314	2	US-08-811-949-48
11	75.4	4.5	1955	2	US-08-883-795A-39
12	75.4	4.5	2457	6	5344773-1
13	75.4	4.5	2544	4	US-09-703-695A-3
14	75.4	4.5	7360	1	US-08-286-740-1
15	75.4	4.5	7360	5	PT-US95-09596-1
16	73.8	4.4	329	1	US-08-148-910-2
17	73.8	4.4	329	1	US-08-448-937A-2
18	73.8	4.4	329	1	US-08-448-937A-13
19	73.8	4.4	329	1	US-08-448-937A-13
20	73.8	4.4	1065	2	US-08-811-949-60
21	73.8	4.4	1068	1	US-08-137-116-2
22	73.8	4.4	1068	1	US-08-427-640-7
23	73.8	4.4	1068	1	US-08-811-949-44
24	73.8	4.4	1068	2	US-08-811-949-46
25	73.8	4.4	1068	2	US-08-811-949-52
26	73.8	4.4	1068	2	US-08-811-949-58
27	73.8	4.4	1068	6	5223256-3

28	73.8	4.4	1163	2	US-08-558-269-5
29	73.8	4.4	1163	3	US-09-410-882-5
30	73.8	4.4	1170	2	US-08-811-949-64
31	73.8	4.4	1170	2	US-08-811-949-66
32	73.8	4.4	1314	2	US-08-811-949-50
33	73.8	4.4	1314	2	US-08-811-949-54
34	73.8	4.4	1314	2	US-08-811-949-56
35	73.8	4.4	1419	2	US-08-811-949-62
36	73.8	4.4	1738	6	5200340-1
37	73.8	4.4	1848	3	US-08-814-412-10
38	73.8	4.4	1974	2	US-08-811-949-38
39	73.8	4.4	2101	2	US-08-811-949-42
40	73.8	4.4	2162	1	US-08-119-512-3
41	73.8	4.4	2162	1	US-08-488-015B-3
42	73.8	4.4	2162	1	US-08-488-015B-25
43	67.8	4.0	1724	6	5200340-5
44	67.8	4.0	2497	6	5185259-2
45	66.4	3.9	2544	3	US-09-518-046-3

ALIGNMENTS

RESULT 1  
US-08-148-910-14  
; Sequence 14, Application US/08148910  
; Patent No. 5466593  
; GENERAL INFORMATION:  
; APPLICANT: Takeshi SHIMOMURA et al.  
; TITLE OF INVENTION: NO. 5466593el Protein and Gene Encoding Said Protein  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch,  
; MEDIUM TYPE: 500 Kb Storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; FILING DATE: No. 5466593ember 5, 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-8850  
; TELEFAX: 202-371-8856  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2033 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; IMMEDIATE SOURCE:  
; LIBRARY: Pre-made Lambda phage Library,  
; LIBRARY: human liver(49, male) cDNA Library (Stratagene)  
US-08-148-910-14



QY 529 GCCTGTCCGACCAAGTTCAAGGGGAAATTCGTGAATAGTTCTGATGA---CTCCTAT 585  
 Db 802 GCCTGTCCGACCAAGTTCTGAGCGCTCTGCAATCATGAGCCTGATGAGCGCTGCTTC 861  
 QY 586 GTTGGCGATGCTACTCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCTGC 645  
 Db 862 TTGGGGAACGGCACTGGGTACCGTGGGTGGGCGGACACCTCAGCTCGGGCCCTCAGCTGC 921  
 QY 646 CTTTACTGGAACTCCCACTCTCTTCCAGAGAAATTTACAACATGTTTATGAGGATGCT 705  
 Db 922 CTGGCTTGGAACTCCGATCTGCTTACAGAGAGTGCACGTGGAGCTCGTGGCGCGCG 981  
 QY 706 GAAACCAATGGATTTGGGGAACACAATTTCTGAGAAACCAAGTGGGACGAAAGCCC 765  
 Db 982 GCCCTGTGGGCTTGGGCGGCTTCTGCTTACCGGAAATCGGACAAATGACGAGGCCCC 1041  
 QY 766 TGGTGTCTTATTAAGTTTACCAATGCAAGGTGAATGGGAATCTGATGCTCAGCC 825  
 Db 1042 TGGTGTCTACGT---GGTGAAGGACGCGCTCTCTTGGGAGTACTGCGGCTGGAGGCC 1098  
 QY 826 TGCTAGCCCAAGGAGTGGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTT 885  
 Db 1099 TGGCAATCCCT-----CACGAGATCCACTGTACCGGATCTCTTGGGAGCCCTG 1149  
 QY 886 CCGGGTTTGAATCTCTGTGGAAGACTGAGATAGAGAGGAAATCAAGAGAAATCTAT 945  
 Db 1150 CTTGAGCAGCTTCCCGGGCGGCGAGCTTCCGCGAGGAGGACAGAGAGAGAGCTTC 1209  
 QY 946 GGAGGCTTTAAGACACGCGGGCAAGCAACCATGGAGGCTCCCTCCAGTCTCTGCTG 1005  
 Db 1210 CTGGGCGCACATATCATCGGGGCTCTCTCTGCTGCGGCTGCGACCCCTG----- 1262  
 QY 1006 CCTGTGACCATCTCATGTCGCCCGGCGACCTTCTGTGTGGGGCGCTGATCCACCCCTGC 1065  
 Db 1263 --GCTGGCGGCACTTACATCGGGAACAGCTTCTGCGCGGAGCCTTGTTCCACCTGC 1320  
 QY 1066 TGGGTGCTCACTGTGCTCCACTGACCGACGACATAAAACACGACATCTAA-----AGGTG 1119  
 Db 1321 TGGGTGCTGCTGCGCGCTTCTGCTTCTCCACAGCCCGGCGGACAGGCTCTCGGTG 1380  
 QY 1120 GTGCTAGGGGACAGGACTGAGAAAGAAATTTATGAGCAGAGTTTAGGGTGGAG 1179  
 Db 1381 GTGCTGGGCGACACTTCTTCAACCGCACGACGAGTGCAGCAGACCTTCGGCATCGAG 1440  
 QY 1180 AAGATATTCAAGTACAGCCACTTACAATGAAGAGATGAGATTCGCCCAATGATTTGCA 1239  
 Db 1441 AAGTACATCCGTACACCTGTACTCGGTTCACACCCCGGACCCAC---GACCTGCTC 1497  
 QY 1240 TTGCTCAAGTTAAAGCAGTGGTCACTGTGCTCTAGATCCAAATAGTGAAGACT 1299  
 Db 1498 CTGATCCGGCTGAGAAGAAAGGAGCCGCTGTGCCACACGCTCGCAGTTCGTGAGGCC 1557  
 QY 1300 GTGTGCTTGCCTGAT-----GGGTCTTCTTCCCTCTGGAGTGAAGTGCACATCTCTGGC 1353  
 Db 1558 ATCTGCTGCCGAGCCCGGAGCAGCTTCCCGCAGGACACAGTGCAGATTTGGGGC 1617  
 QY 1354 TGGGG-----TGTTACGAAACAGAGAAAGGGTCCGCGAGCTCTCTGGATGCCAAAGTC 1407  
 Db 1618 TGGGGCCACTTGGATGAGAAAGTGAAGGCTTACTCCAGCTCTCCCTCGGGAGCCCTGTGTC 1677  
 QY 1408 AAGCTGATTTGCAACACTTTGTGCACTCCCGCAACTCTATGACCAATGATGATGAC 1467  
 Db 1678 CCCCTGGTCCGCAACCAAGTGCAGCAGCTTGGAGTCTAGCGGCGGACATCAGCCCC 1737  
 QY 1468 AGTATGATCTGTGAGGAAATCTTTCAGAAACCTTGGGCAAGACCTTGGCAGGCTGACTCT 1527  
 Db 1738 AACATGCTCTGTGCGGGTACTT---CGACTGCAAGTCCGAGCGCTTCCGAGGGGACTCA 1794  
 QY 1528 GGAGGCGCCCTGACTGTGAGAGGAGCGCACTTACTAGTCTATGGGATAGTGGAGTGG 1587  
 Db 1795 GGGGGGCCCCCTGGCCTGCGAGAGAAACGGCGTGGCTTACTCTACGCGATCATCAGCTGG 1854

QY 1588 GGCCTGGAGTGGG-----AAGAGCCAGGGCTTACACCCCAAGTTTACCAATTTCTG 1641  
 Db 1855 GGTGACGCTCGGGCGGCTCCACAGCCGGGGTCTTACACCCCGTGGCCACTATGTG 1914  
 QY 1642 AATTGGATCAAAAGCCACCAT 1661  
 Db 1915 GACTGGATCAACGACCGGAT 1934

RESULT 3  
 US-08-148-910-3  
 ; Sequence 3, Application US/08148910  
 ; Patent No. 546593  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takeishi SHIMOMURA et al.  
 ; TITLE OF INVENTION: No. 546593el Protein and Gene Encoding Said Protein  
 ; NUMBER OF SEQUENCES: 14  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Wenderoth, Lind & Ponack  
 ; STREET: 805 Fifteenth Street, N.W., #700  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: U.S.A.  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 5.25 inch,  
 ; MEDIUM TYPE: 500 Kb Storage  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: Wordperfect  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/148,910  
 ; FILING DATE: No. 546593ember 5, 1993  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Warren M. Cheek, Jr.  
 ; REGISTRATION NUMBER: 33,367  
 ; REFERENCE/DOCKET NUMBER:  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-371-8850  
 ; TELEFAX: 202-371-8856  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 970 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: human  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: Pre-made Lambda phage Library, human liver (49, male)  
 ; LIBRARY: cDNA Library (Stratagene)  
 ; US-08-148-910-3

Query Match 5.7%; Score 96.2; DB 1; Length 970;  
 Best Local Similarity 53.0%; Pred. No. 8.8e-19;  
 Matches 359; Conservative 0; Mismatches 288; Indels 30; Gaps 6;

QY 1009 CTGACCATCTCCATGCCCGGAGGCGGCTTCTGTGTGGGCGCTGATCCACCCCTGCTGG 1068  
 Db 201 CTGGCGCCCATCTACATCGGGGACAGCTTCTGCGCGGAGCCTGTGTCACACCTGCTGG 260  
 QY 1069 GTGCTCACTGTCGCCCTGACCGACATATAAAACACGACA-----TCTAAGGTGGTG 1122  
 Db 261 GTGGTGTGCGCGCGGCTCTCTTCTCCACAGCCCCCCCCAGGACAGCGTCTCCGTGGTG 320  
 QY 1123 CTAGGGACACGAGGACCTGAAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAAG 1182

321 CTGGCCAGCAGCTTCTTCAACGACGACGAGCGTGCAGCAGACCTTCGGCATCAGAAG 380  
 1183 ATATTCAAGTACAGCAGCTACATGAAGAGATGAGATTCCTCCCAATGATATTCATTTG 1242  
 381 TACATCCGCTACACCTGTACTCGGTGTTCAACCCCA---GGACACAGACCTCTCTCTG 437  
 1243 CTCAGTTAAGCCAGTGGTGTCTGTCTGTAGATTCCTCAATATCCTGAAGACTGTG 1302  
 438 ATCCGGCTGAAGAAAGGAGGACGCTGTGCACACGCTGCGAGTTCTGTGAGCCCATC 497  
 1303 TGCTTCCCTGATG-----GGTCTTTCCCTCTGGGAGTGAAGTGCACATCTCTGGCTGG 1356  
 498 TGCTTCCCTGATG-----GGTCTTTCCCTCTGGGAGTGAAGTGCACATCTCTGGCTGG 1356  
 1357 GG-----TGTACAGAAACAGAAAGGCTCCGCGAGCTCTCTGTGATCCCAAGTCAAG 1410  
 558 GGCCACTTGTGATGAGAAAGTGAAGGCTATCTCCAGCTCTCTGGGAGGCTCTGGTCC 617  
 1411 CTGATTGCCAAGCTTTGTGCAACTCCGCAACTCTATGACCAATGATGATGACAGT 1470  
 618 CTGGTCCGACACCAAGTGCAGCAGCTGTAGGTTCTACGGGCGGACATCAGCCCAAC 677  
 1471 ATGATCTGTGCAGAAATCTTCAAGAACTTCAAGAACTTCAAGAACTTCAAGAACTT 1530  
 678 ATGCTCTGTGCGGCTACTT---CGACTGCAAGTCCGAGCGCTGCCAGGGGACTCAGG 734  
 1531 GGCCCTTACCTGTGAGAAAGGAGCGGACCTACTACTGTATGGGATGATGAGTGGGCG 1590  
 735 GGCCCTTACCTGTGAGAAAGGAGCGGCTACTACTGTATGGGATGATGAGTGGGCG 794  
 1591 CTGAGTGTGGG-----AAGAGCCAGGGGTCTACACCAAGTTCACCAATTCCTGAAT 1644  
 795 GACGGTGTGGGCGGCTTCCACAGCGGGGTCTACACCGCGTGCAGCACTATGTGGAC 854  
 1645 TGGATCAAGCCACCAT 1661  
 855 TGGATCAAGCCGAT 871

# RESULT 4

US-08-448-937A-3  
 ; Sequence 3, Application US/08448937A  
 ; Patent No. 5677164  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takeshi SHIMOMURA et al.  
 ; TITLE OF INVENTION: No. 5677164el Protein and Gene Encoding Said Protein  
 ; NUMBER OF SEQUENCES: 14  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Wenderoth, Lind & Ponack  
 ; STREET: 805 Fifteenth Street, N.W., #700  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: U.S.A.  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 5.25 inch,  
 ; MEDIUM TYPE: 500 Kb Storage  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: Wordperfect  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/448,937A  
 ; FILING DATE: May 24, 1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/148,910  
 ; FILING DATE: No. 5677164ember 5, 1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Warren M. Cheek, Jr.  
 ; REGISTRATION NUMBER: 33,367  
 ; REFERENCE/DOCKET NUMBER:  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-371-8850

TELEFAX: 202-371-8856  
 TELEX:  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 970 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: human  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: Pre-made lambda phage Library, human liver (49, male)  
 ; LIBRARY: cDNA Library (Stratagene)  
 ; US-08-448-937A-3

Query Match 5.7%; Score 96.2; DB 1; Length 970;  
 Best Local Similarity 53.0%; Pred. No. 8.8e-19;  
 Matches 359; Conservative 0; Mismatches 288; Indels 30; Gaps 6;  
 QY 1009 CTGACCATCTCCATGCCCCAGGGCCACTTCTGTGTGGGGCGGTGATCCACCCCTGCTGG 1068  
 DB 201 CTGGCCGCCATCTACATCGGGGACAGCTTCTGCGCGGGAGCCTGTGTCACACCTGCTGG 260  
 QY 1069 GTGCTCACTGTGCCCCACTGCACCGACATAAAAAACACAGACA-----TCTAAAGGTGGT 1122  
 DB 261 GTGTGTGGCGCGCCACTGCTTCTCCACAGCCCCCCCCAGGGACAGCGTCTCGTGGTGG 320  
 QY 1123 CTAGGGACACGAGACCTGGAAGAAAGAAATTTCTAGAGCAGAGCTTTAGGTGGAGAAG 1182  
 DB 321 CTGGCCGAGCACTTCTTCAACCCGACGACGAGCTGTGACGAGACCTTTGGCATCGAGAAG 380  
 QY 1183 ATATTCAAGTACAGCCACTACATGAAGAGATGAGATTCCTCCACCAATGATATTGCTATG 1242  
 DB 381 TACATCCGTTACACCTGTACTCGGTGTTCAACCCCA---GCGACCAAGACCTCGTCTG 437  
 QY 1243 CTCAAGTTAAAGCCAGTGGATGGTCACTGTGCTCTAGATTCCTAGATTCCTAGATTCCTAG 1302  
 DB 438 ATCCGGCTGAAGAAAGGAGGAGCGCTGTGTCACACGCTCGCAGTTCTGTGAGCCCATC 497  
 QY 1357 GG-----TGTTACAGAAACAGGAAAGGTTCCCGCAGCTCTCTGGATGCAAAAGTCAAG 1410  
 DB 558 GGCCACTTGGATGAGAACGCTGAGCGGCTACTTCCAGCTCTCTGGGAGGCGCTGGTCCCC 617  
 QY 1411 CTGATTGCCAAGCTTTGTGCAACTCCGCGCAACTCTATGACCAATGATGATGACAGT 1470  
 DB 618 CTGGTCCGCGACCAAGTGTGAGCAGCGCTGTAGGTTCTACGGCGCGGACATCAGCCCCAAC 677  
 QY 1471 ATGATCTGTGAGGAATCTTCAAGAACTTCAAGAACTTCAAGAACTTCAAGAACTTCTGGA 1530  
 DB 678 ATGCTCTGTGCGGCTACTT---CGACTGCAAGTCCGAGCGCTGCCAGGGGACTCAGG 734  
 QY 1531 GGCCCTTACCTGTGAGAAAGGAGCGGACCTACTACTGTGATGATGAGTGGGCG 1590  
 DB 735 GGCCCTTACCTGTGAGAAAGGAGCGGCTACTACTGTGATGATGAGTGGGCG 794  
 QY 1591 CTGGAGTGTGGG-----AAGAGCCAGGGGTCTACACCAAGTTCACCAATTCCTGAAT 1644  
 DB 795 GACGGTGTGGGCGGCTTCCACAGCGGGGTCTACACCGCGTGCAGCACTATGTGGAC 854  
 QY 1645 TGGATCAAGCCACCAT 1661  
 DB 855 TGGATCAAGCCGAT 871

# RESULT 5

US-08-427-640-1  
 ; Sequence 1, Application US/08427640  
 ; Patent No. 5658798



GENERAL INFORMATION:  
APPLICANT: Berg et al.  
TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disorder  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: IN  
COUNTRY: U.S.A.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/427,640  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: 07/689,410  
FILING DATE: 22 APRIL 1991  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1065 base pairs  
TYPE: Nucleic acid  
STRANDEDNESS: Double  
TOPOLOGY: Linear  
MOLECULE TYPE: DNA  
US-08-427-640-1

Query Match 4.5%; Score 75.4; DB 1; Length 1065;  
Best Local Similarity 50.8%; Pred. No. 1.7e-12;  
Matches 342; Conservative 0; Mismatches 286; Indels 45; Gaps 5;  
QY 1028 AGGCCACTTCTGTGGGGCGCTGATCCACCCCTGTGGGTGCTCACTGTGCCCACT 1087  
DB 392 AGCGTTCTCTGCGGGGCATCTCATAGCTCTCTGCGGCTCTCTGCGGCCACT 451  
QY 1088 GCACCGA-----CATAAAAACCGACATCTAAAGTGTGTAGGGACCGAGCTGA 1141  
DB 452 GCTTCCAGGAGAGGTTTCCGCCACCACTGACGTGATCTGGCGAGAACATACCGG 511  
QY 1142 AGAAGAAGAAATTCATGACGAGCTTTAGGTTGAGAGATATTCAGTACACCCACT 1201  
DB 512 TGGTCCCTGCGAGGAGCGAATTTGAAGTCGAAAAATACATTTGTCATAAGGAAT 571  
QY 1202 ACAATGAAAGAGATGAGATTCCCAATGATATTGCAATTGCTCAAGTTAAAGCCAGTGG 1261  
DB 572 TCGAT-----GATGACACTTACGACATGACATTCGCTGCTGCAGCTGAATCGGATT 625  
QY 1262 ATGGTCACTGTGCTTAGAATCCAAATAGTGAAGCTGTGTGCTTGCTGATGGTCT 1321  
DB 626 CGTCCGCTGTGCCAGGAGCAGCGTGTCCGCACTGTGTGCTTCCCGCGCGGACC 685  
QY 1322 TTCCT-----CTGGGAGTGAAGTGCACATCTCTGGTGGGTGTTACAGAA 1369  
DB 686 TGCAGTCCGAGTGCAGCGAGTGTGAGCTCTCCGGCTACGGCAAGCATGAGCCTTGT 745  
QY 1370 CAGGAAAGGTCCTCCAGCTCTCTGATGCCAAAGTCAAGCTGATTCGCAACATTTGT 1429  
DB 746 CTCCTTTCTATTCCGAGCGGCTGAGGAGGCTCATGTCAAGCTGTACCCATCCAGCGCT 805  
QY 1430 GCAACTCCGCGCAACTCTATGACCATGATGATGACAGATGATGCTGTGAGGAAATC 1489  
DB 806 GCATCATCAACATTTACTTAACAGAACAGTACCAGCAACATGCTGTGTGTGAGACA 865  
QY 1490 TTCAGAAACCTGG-----GCAGACACCTGCCAGGCTGACTGTGAGGACC 1534  
DB 866 CTCGAGGCGCGGCGCCAGGCAAACTTGACGACGCTGCCAGGCGATTCGGGAGGCC 925  
QY 1535 CCCTGACCTGTGAGAGGAGCGGACCTACTACGCTCTATGGATAGTGAAGTGGGCGCTGG 1594

DB 926 CCTGTGTGTCTGAACGATGCGCGCATCTTGGTGGGCATCATCAGCTGGGCGCTGG 985  
QY 1595 AGTGTGGGAGAGG-----CCAGGGTCTACACCCAGTTACCAAAATTCCTGAATTGGA 1648  
DB 986 GCTGTGACAGAGAGGATGTCGCGGTGTGTACCAAGGTTTACCAACTACCTAGACTGA 1045  
QY 1649 TCAAAGCCACCAT 1661  
DB 1046 TTCTGTGACACAT 1058  
RESULT 6  
US-08-427-640-5  
Sequence 5, Application US/08427640  
Patent No. 5658788  
GENERAL INFORMATION:  
APPLICANT: Berg et al.  
TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disorder  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: IN  
COUNTRY: U.S.A.  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/427,640  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: 07/689,410  
FILING DATE: 22 APRIL 1991  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1065 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-427-840-5

Query Match 4.5%; Score 75.4; DB 1; Length 1065;  
Best Local Similarity 50.8%; Pred. No. 1.7e-12;  
Matches 342; Conservative 0; Mismatches 286; Indels 45; Gaps 5;  
QY 1028 AGGCCACTTCTGTGGGGCGCTGATCCACCCCTGTGGGTGCTCACTGTGCCCACT 1087  
DB 392 AGCGTTCTCTGCGGGGCATCTCATAGCTCTCTGCGGCTCTCTGCGGCCACT 451  
QY 1088 GCACCGA-----CATAAAAACCGACATCTAAAGTGTGTAGGGACCGAGCTGA 1141  
DB 452 GCTTCCAGGAGAGGTTTCCGCCACCACTGACGTGATCTGGCGAGAACATACCGG 511  
QY 1142 AGAAGAAGAAATTCATGACGAGCTTTAGGTTGAGAGATATTCAGTACAGCCACT 1201  
DB 512 TGGTCCCTGCGAGGAGCGAATTTGAAGTCGAAAAATACATTTGTCATAAGGAAT 571  
QY 1202 ACAATGAAAGAGATGAGATTCCCAATGATATTGCAATTGCTCAAGTTAAAGCCAGTGG 1261  
DB 572 TCGAT-----GATGACACTTACGACATGACATTCGCTGCTGCAGCTGAATCGGATT 625  
QY 1262 ATGGTCACTGTGCTTAGAATCCAAATAGTGAAGCTGTGTGCTTGCTGATGGTCTCT 1321  
DB 626 CGTCCGCTGTGCCAGGAGCAGCGTGTCCGCACTGTGTGCTTCCCGCGCGGACC 685  
QY 1322 TTCCT-----CTGGGAGTGAAGTGCACATCTCTGGTGGGTGTTACAGAA 1369



```

QY 1028 AGGGCCACTTCTGTGGGGCGGTGATCCACCCCTGCTGGGTGCTCACTGCTGCCACT 1087
DB 461 AGCGTTCTGTGGGGGCGGTACTCATCAGCTCTCTGTGGATCTCTGTGGCGCCACT 520
QY 1088 GCACCGA-----CATAAACACAGACATCTAAAGGTGGTCTAGGGGACACGACCTGA 1141
DB 521 GCTTCAGGAGAGGTTTCGGGCCACACCTGACGGTGATCTTGGGAGAACATACCGGG 580
QY 1142 AGAAGAGAGAAATTCATGAGCAGAGCTTTAGGGTGGAGAGATATCAAGTACAGCCACT 1201
DB 581 TGGTCCCTGGCGAGAGAGAGAGAAATTTGAATCGAATAATACATTTGTCTCAAGGAAT 640
QY 1202 ACAATGAAGAGATGAGATTCGCCCAATGATATTCATGCTCAAGTTAAAGCCAGTGG 1261
DB 641 TCGAT-----GATGACATTTACGACAAATGACATTTGGCTGCTGCAGCTGAAATCGGAT 694
QY 1262 ATGGTCACGTGCTCTAGAAATCCAAATACGTTGAAGACTGTGTGCTTCCCTGATGGGTCCT 1321
DB 695 CGTCCCGCTGTGCCAGAGAGACGCTGGTCCGCACTGTGTGCTTCCCGCGCGACC 754
QY 1322 TTCCCT-----CTGGAGTGTAGTGCCACATCTCTGGCTGGGGTGTACAGAA 1369
DB 755 TGCAGCTGCCGAGTGGAGCGAGTGTAGCTCTCCGGCTACGGCAAGCATGAGGCTTGT 814
QY 1370 CAGGAAAGGGTCCCGCAGCTCTGGATGCCAAAGTCAAGCTGATTTGCCAACACTTTGT 1429
DB 815 CTCCTTTCTATTGGAGCGGCTGAGGAGGCTCATGTACAGCTGTACCCATCCAGCGCT 874
QY 1430 GCAATCCCGCCAACTCTATGACCAATGATGATGACAGATGATGCTGTGAGGAAATC 1489
DB 875 GCACATCAACATTTACTTAACAGACAGTCAACGACACATCTGTGTGCTGGAGACA 934
QY 1490 TTCAAGAACCTGG-----GCAAGACACTGCCAGGGTGTACTCTGAGGGCC 1534
DB 935 CTCGAGCGGGCGGCCAGGCAAACTTGCACGACCTGCCAGGCGGATTCGGAGGCC 994
QY 1535 CCCTGACCTGTGAGAGAGCGGACCTACTACTGCTATGAGATAGTGTGAGCTGGGCTGG 1594
DB 995 CCCTGGTGTGTGAACGATGCCGCGATGCTTTGGTGGGCATCATCAGCTGGGCGCTGG 1054
QY 1595 AGTGTGGGAGAGG-----CCAGGGGTCTACACCAAGTTACCAAAATTCCTGAATTGGA 1648
DB 1055 GCTGTGGACAGAGGATGTCCGGGTGTGTACACCAAGGTTACCAACTACTAGACTGGA 1114
QY 1649 TCAAGGCCACCAT 1661
DB 1115 TTCGTGACAACT 1127

```

RESULT 9

```

; US-09-618-869-9
; Sequence 9, Application US/09618869
; Patent No. 6455279
; GENERAL INFORMATION:
; APPLICANT: Ambrosius, Dorthie
; APPLICANT: Rudolph, Rainer
; APPLICANT: Schaeffner, Joerg
; APPLICANT: Schwarz, Elisabeth
; TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF NATURALLY FOLDED AND
; SECRETED PROTEINS BY CO-SECRETION OF MOLECULAR
; CHAPERONES
; FILE REFERENCE: 20381
; CURRENT APPLICATION NUMBER: US/09/618,869
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: EP91114811.5
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: (1)...(1137)
US-09-618-869-9

Query Match      4.5%; Score 75.4; DB 4; Length 1137;
Best Local Similarity 50.8%; Pred. No. 1.8e-12;
Matches 342; Conservative 0; Mismatches 286; Indels 45; Gaps 5;

QY 1028 AGGGCCACTTCTGTGGGGCGGTGATCCACCCCTGCTGGGTGCTCACTGCTGCCACT 1087
DB 461 AGCGTTCTGTGGGGGCGGTACTCATCAGCTCTCTGTGGATCTCTGTGGCGCCACT 520
QY 1088 GCACCGA-----CATAAACACAGACATCTAAAGGTGGTCTAGGGGACACGACCTGA 1141
DB 521 GCTTCAGGAGAGGTTTCGGGCCACACCTGACGGTGATCTTGGGAGAACATACCGGG 580
QY 1142 AGAAGAGAGAAATTCATGAGCAGAGCTTTAGGGTGGAGAGATATCAAGTACAGCCACT 1201
DB 581 TGGTCCCTGGCGAGAGAGAGAGAAATTTGAATCGAATAATACATTTGTCTCAAGGAAT 640
QY 1202 ACAATGAAGAGATGAGATTCGCCCAATGATATTCATGCTCAAGTTAAAGCCAGTGG 1261
DB 641 TCGAT-----GATGACATTTACGACAAATGACATTTGGCTGCTGCAGCTGAAATCGGAT 694
QY 1262 ATGGTCACGTGCTCTAGAAATCCAAATACGTTGAAGACTGTGTGCTTCCCTGATGGGTCCT 1321
DB 695 CGTCCCGCTGTGCCAGAGAGACGCTGGTCCGCACTGTGTGCTTCCCGCGCGACC 754
QY 1322 TTCCCT-----CTGGAGTGTAGTGCCACATCTCTGGCTGGGGTGTACAGAA 1369
DB 755 TGCAGCTGCCGAGTGGAGCGAGTGTAGCTCTCCGGCTACGGCAAGCATGAGGCTTGT 814
QY 1370 CAGGAAAGGGTCCCGCAGCTCTGGATGCCAAAGTCAAGCTGATTTGCCAACACTTTGT 1429
DB 815 CTCCTTTCTATTGGAGCGGCTGAGGAGGCTCATGTACAGCTGTACCCATCCAGCGCT 874
QY 1430 GCAATCCCGCCAACTCTATGACCAATGATGATGACAGATGATGCTGTGAGGAAATC 1489
DB 875 GCACATCAACATTTACTTAACAGACAGTCAACGACACATCTGTGTGCTGGAGACA 934
QY 1490 TTCAAGAACCTGG-----GCAAGACACTGCCAGGGTGTACTCTGAGGGCC 1534
DB 935 CTCGAGCGGGCGGCCAGGCAAACTTGCACGACCTGCCAGGCGGATTCGGAGGCC 994
QY 1535 CCCTGACCTGTGAGAGAGCGGACCTACTACTGCTATGAGATAGTGTGAGCTGGGCTGG 1594
DB 995 CCCTGGTGTGTGAACGATGCCGCGATGCTTTGGTGGGCATCATCAGCTGGGCGCTGG 1054
QY 1595 AGTGTGGGAGAGG-----CCAGGGGTCTACACCAAGTTACCAAAATTCCTGAATTGGA 1648
DB 1055 GCTGTGGACAGAGGATGTCCGGGTGTGTACACCAAGGTTACCAACTACTAGACTGGA 1114
QY 1649 TCAAGGCCACCAT 1661
DB 1115 TTCGTGACAACT 1127

```

RESULT 10

```

; US-08-611-949-48
; Sequence 48, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE FLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.

```

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 05-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 18-966-0  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1314 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1311  
US-08-811-949-48

Query Match  
Best Local Similarity 50.8%; Pred. No. 1.9e-12; Indels 45; Gaps 5;  
Matches 342; Conservative 0; Mismatches 286;

QY 1028 AGGCCACTCTGTGGGGGGGGTATCCACCCCTGCTGGGTGCTCACTGTGCCCACT 1087  
Db 638 AGCGTTCCTGTGGGGGGGATCTATCAGCTCTCTGCTGATTCCTCTGCGCCACT 697  
QY 1088 GCACCGA-----CATAAAAACCAAGCATCTAAAGTGTGCTAGGGACCAAGCACTGA 1141  
Db 698 GCTTCCAGGAGAGGTTTCCGCCCCACCACTGACGGTGTCTTTGGGCGAACAATACCGGG 757  
QY 1142 AGAAGAAGATTTTCATGACGAGAGCTTTAGGGTGGAGAGATATTCAGTACAGCCACT 1201  
Db 758 TGGTCCCTGGCGAGAGAGAGAGAAATTTGAAGTCGAAATATCATGTCCATAGGAAT 817  
QY 1202 ACAATGAAGAGATGAGATTCACCAATGATATGCTCAATGCTCAAGTTAAAGCCAGTGG 1261  
Db 818 TCGAT-----GATGACACTTACCAATGACATTCGCTGCTGCTGCTGCTGCTGCTGCT 871  
QY 1262 ATGTCACCTGTCTAGATCCAAATACGTAAGAGCTGTGTGCTTCCCTGATGGGTCT 1321  
Db 872 GGTCCCGCTGTGCCAGAGAGAGAGCGGTGGTCCGCTATGTGCTTCCCGCGGACC 931  
QY 1322 TTCCCT-----CTGGAGTGTAGTCCACATCTCTGTGGTGGGTGTACAGAA 1369  
Db 932 TGCAGTCCCGACTTGGACGGAGTGTGAGCTCTCCGCTACCGCAAGCATGAGCCCTGT 991  
QY 1370 CAGGAAGAGTCCCGCAGCTCTGATGCCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1429  
Db 992 CTCCTTTCTATTCGAGCGGTGAAGAGGCTCATGTCAGACTGTACCCATCCAGCCGCT 1051  
QY 1430 GCACTCCCGCACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1489  
Db 1052 GCACATCAACATTTACTTAACAGACAGTACAGCAACATGCTGTGTGCTGAGACA 1111  
QY 1490 TTCAGAAACCTGG-----GCAGACACCTGCCAGGTGACTCTGAGGCC 1534  
Db 1112 CTCGAGCGCGGGGCCCCAGGCAAACTTGCAGAGCGCTGCCAGGGGATTCGGGAGGCC 1171

QY 1535 CCTGACCTGTGAGAAAGGACGGACCTACTAGTCTATGGGATAGTGTGAGCTGGGCGCTGG 1594  
Db 1172 CCTGGTGTCTGAAACGATGGCCGCATCATTGGTGGGCATCATCAGCTGGGCGCTGG 1231  
QY 1595 AGTGTGGGAGAGG-----CCAGGGGTCTACACCAAGTACCAAAATTCCTGAATGGA 1648  
Db 1232 GCTGTGACAGAGGATGTCCTCCGGGTGTGTACACAAAGGTTACCAACTACCTAGACTGA 1291  
QY 1649 TCAAGGCCACCAT 1661  
Db 1292 TTGCTGACACAT 1304

RESULT 11  
US-08-883-795A-39  
Sequence 39, Application US/08893795A  
Patent No. 5985607  
GENERAL INFORMATION:  
APPLICANT: Delcove, Genevieve  
APPLICANT: Awang, Gregor  
TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 27-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Micheline  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 7841-062  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1955 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-883-795A-39

Query Match  
Best Local Similarity 50.8%; Pred. No. 2.4e-12; Indels 45; Gaps 5;  
Matches 342; Conservative 0; Mismatches 286;

QY 1028 AGGCCACTCTGTGGTGGGGGGTGTATCCACCCCTGCTGGGTGCTCACTGTGCCCACT 1087  
Db 1103 AGCGTTCCTGTGGGGGGGATCTATCAGCTCTCTGCTGATTCCTCTGCGCCCACT 1162  
QY 1088 GCACCGA-----CATAAAAACCAAGCATCTAAAGTGTGCTAGGGACCAAGCACTGA 1141  
Db 1163 GCTTCCAGGAGAGGTTTCCGCCCCACCACTGACCGTGTATCTTGGGCGAGACATACCGGG 1222  
QY 1142 AGAAGAAGATTTTCATGACGAGAGCTTTAGGGTGGAGAGATATTCAGGTACAGCCACT 1201  
Db 1223 TGTCTCCTGGCGAGGAGGAGAGAAATTTGAAGTCGAAAAATACATTTGTCATAGGAAT 1282





ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/09576  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/286740  
FILING DATE: 05-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 798PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7360 bases  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
PCT-US95-09576-1

Query Match 4.5%; Score 75.4; DB 5; Length 7360;  
Best Local Similarity 50.8%; Pred. No. 5e-12;  
Matches 342; Conservative 0; Mismatches 286; Indels 45; Gaps 5;

QY 1028 AGGGCCACTTCGTGGTGGGGCGCTGATCCACCCTGTGGGTGCTCACTGCTGCCCACT 1087  
DB |||||  
QY 2634 AGCGGTTCCTGTGGGGGGCATACATCAGCTCCTGTGGATTCTCTCGCGCCCACT 2693  
DB |||||  
QY 1088 GCACCGA-----CATAAAAACAGACATCTAAAGGTGGTGTAGGGGACCAGGACCTGA 1141  
DB |||||  
QY 2694 GCTTCCAGGAGAGGTTCCGCCCCACCACCTGACGGTGATCTTGGGCAGAACATACCGGG 2753  
DB |||||  
QY 1142 AGAAGAGAAATTTCATGAGCAGAGCTTTAGGTGGAGAGATATTCAAGTACAGCCACT 1201  
DB |||||  
QY 2754 TGGTCCCTGGCGAGGAGGAGCAGAAATTTGAAGTCGAAAAATACATTTGTCATAGGAAT 2813  
DB |||||  
QY 1202 ACAATGAAGAGATGAGATTCCCAACATGATTTGCTCAAGTTAAAGCCAGTGG 1261  
DB |||||  
QY 2814 TCGAT-----GATGACATTAGCACATGACATTGCGCTGCTGAGCTGAAATCGGATT 2867  
DB |||||  
QY 1262 ATGGTCACGTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCTGATGGGTCT 1321  
DB |||||  
QY 2868 CGTCCCGCTGTGCCAGGAGAGCAGCGTGGTCCGCACTGTGTGCTTCCCGCGCGACC 2927  
DB |||||  
QY 1322 TTCCCT-----CTGGGAGTGAAGTCCACATCTCTGGCTGGGGTGTACAGAA 1369  
DB |||||  
QY 2928 TGAGCTGCCGAGCTGGAGGAGTGTGAGCTCTCGGCTACGGCAAGCATGAGGCTTGT 2987  
DB |||||  
QY 1370 CAGGAAAGGTTCCGCCAGCTCTGGATGCCAAAGTCAAGCTGATGCCAACACTTTGT 1429  
DB |||||  
QY 2988 CTCCTTTCTATTCCGAGCGGCTGAAGGAGGCTCATGTGACTGTACCCATCCAGCCGT 3047  
DB |||||  
QY 1430 GCAACTCCGCCAATCTATGACCAATGATTTGATGACAGATGATCTGTGAGGAATC 1489  
DB |||||  
QY 3048 GCACATCAACAATTACTTAACAGAACAGTACCAGCAACATGCTGTGTGTGGAGACA 3107  
DB |||||  
QY 1490 TTCAGAAACCTGG-----GCAAGACACTGCCAGGCTGACTCTGGAGGCC 1534  
DB |||||  
QY 3108 CTCGAGCGCGGGCCCCAGGCAAACTTCACAGCGCTGCCAGGGGATTCGGGAGGCC 3167  
DB |||||  
QY 1535 CCTGACCTGTGAGAGGAGCGCACCTACTAGTCTATCGGATAGTAGCTGGGCGCTGG 1594  
DB |||||  
QY 3168 CCTGGTGTGTGTAACGATGGCCGCATGACTTTGGTGGGCATCATCAGCTGGGCGCTGG 3227  
DB |||||

QY 1595 AGTGTGGGAAGAGG-----CCAGGGGTCTACACCAAGTTACCAAAATTCCTGAATTGGA 1648  
DB |||||  
QY 3228 GCTGTGGACAGAAGGATGTCCCGGGTGTGTACACCAAGGTACCAACTACTAGACTGGA 3287  
DB |||||  
QY 1649 TCAAGGCCACCAT 1661  
DB |||||  
QY 3288 TTGCTGACACAT 3300  
DB |||||

Search completed: May 25, 2004, 10:46:15  
Job time : 98.5 secs



**This Page Blank (uspto)**

Sequence 10, Appl  
Sequence 6, Appl  
Sequence 25, Appl  
Sequence 25, Appl  
Sequence 25, Appl  
Sequence 25, Appl  
Sequence 541, App  
Sequence 541, App  
Sequence 7, Appl  
Sequence 5, Appl  
Sequence 540, App  
Sequence 540, App  
Sequence 3, Appl  
Sequence 144, App  
Sequence 19992, A  
Sequence 174961,  
Sequence 174961,  
Sequence 4, Appl  
Sequence 7, Appl  
Sequence 2, Appl  
Sequence 5, Appl  
Sequence 2, Appl  
Sequence 33, Appl  
Sequence 57, Appl  
Sequence 56, Appl  
Sequence 2167, Ap  
Sequence 16, Appl  
Sequence 133, App  
Sequence 14, Appl  
Sequence 17, Appl  
Sequence 131, App

2299 13 US-10-665-216-10  
1689 9 US-09-969-271-6  
4.5 1726 13 US-10-411-037-25  
4.5 1726 13 US-10-411-026-25  
4.5 1726 17 US-10-410-962-25  
4.5 1726 17 US-10-411-049-25  
4.5 2461 13 US-10-342-887-541  
4.5 2461 13 US-10-172-118-541  
4.5 2509 15 US-10-193-656-7  
4.5 2509 9 US-09-969-271-5  
4.5 2519 13 US-10-342-887-540  
4.5 2519 13 US-10-172-118-540  
4.5 2544 15 US-10-443-701-3  
4.5 2641 9 US-09-974-298-144  
4.5 2859 10 US-09-814-353-19992  
4.4 3186778 13 US-10-027-632-174961  
4.4 3186778 16 US-10-027-632-174961  
4.4 1065 10 US-09-987-455-4  
4.4 1065 10 US-09-987-455-7  
4.4 1128 10 US-09-987-455-2  
4.4 1128 10 US-09-987-455-5  
4.4 1996 16 US-10-401-077-2  
4.4 614 9 US-09-879-792-33  
4.3 7218 4.3 7596 15 US-10-004-113-57  
4.1 8064 15 US-10-004-113-56  
4.1 8221 12 US-10-152-319A-2167  
4.0 1078 10 US-09-898-837A-16  
4.0 1146 13 US-10-428-275-133  
4.0 1314 10 US-09-898-837A-14  
4.0 1314 10 US-09-898-837A-17  
4.0 1314 13 US-10-428-275-131

ALIGNMENTS

RESULT 1

US-09-912-559-1  
; Sequence 1, Application US/09912559  
; Patent No. US20020142316A1  
; GENERAL INFORMATION:  
; APPLICANT: ROEMISCH, JUERGEN  
; APPLICANT: STOEHR, HANS-ARNOLD  
; APPLICANT: FRUSSLER, ANNETTE  
; APPLICANT: LANG, WIEGAND  
; APPLICANT: WEIMER, THOMAS  
; APPLICANT: BECKER, MARGRET  
; APPLICANT: NERLICH, CLAUDIA  
; APPLICANT: MUTH-NAUMANN, GUDRUN  
; TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND  
; TITLE OF INVENTION: DETECTION METHODS USING SPECIFIC ANTIBODIES  
; FILE REFERENCE: 08478.1457  
; CURRENT APPLICATION NUMBER: US/09/912,559  
; CURRENT FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: DE 100 36 641.4  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: DE 100 50 040.4  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: DE 100 52 319.6  
; PRIOR FILING DATE: 2000-10-21  
; PRIOR APPLICATION NUMBER: DE 101 18 706.8  
; PRIOR FILING DATE: 2001-04-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1683  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-912-559-1

Query Match 100.0%; Score 1683; DB 9; Length 1683;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 09:00:39 ; Search time 528.5 Seconds  
(without alignments)  
14472.614 Million cell updates/sec

US-09-912-559-1

Perfect score: 1683  
Sequence: 1 atgtttccaggtgtctga.....aaagtgaagtggtctctaa 1683

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2953838 seqs, 2272363821 residues

Total number of hits satisfying chosen parameters: 5907676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1683	100.0	1683	9	US-09-912-559-1
2	1683	100.0	1683	16	Sequence 1, Appli
3	1683	100.0	3088	9	Sequence 1668, Ap
4	1681.4	99.9	1683	16	Sequence 2, Appli
5	1681.4	99.9	1683	16	Sequence 3, Appli
6	1679.8	99.8	1683	9	Sequence 4, Appli
7	1679.8	99.8	1683	16	Sequence 8242, Ap
8	368.4	21.9	428	10	Sequence 17, Appl
9	262.4	15.6	264	13	Sequence 11164, A
10	120.2	7.1	451	9	Sequence 552, App
11	117.6	7.0	2036	9	Sequence 1612, Ap
12	117.6	7.0	2036	9	Sequence 27791, A
13	117	7.0	117	9	Sequence 590, App
14	93.2	5.5	1302	13	

QY 1081 GCCCACTGACCCGACATATAAACCCAGACATCTAAAGGTGGTCTAGGGGACCCAGGACCTG 1140  
DB 1081 GCCCACTGACCCGACATATAAACCCAGACATCTAAAGGTGGTCTAGGGGACCCAGGACCTG 1140  
QY 1141 AAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAGATATTCAGGTACAGCCAC 1200  
DB 1141 AAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGGAGAGATATTCAGGTACAGCCAC 1200  
QY 1201 TACATGAAAGAGATGAGATTCGCCCAATGATTTGCAATGCTCAAGTTAAAGCCAGTG 1260  
DB 1201 TACATGAAAGAGATGAGATTCGCCCAATGATTTGCAATGCTCAAGTTAAAGCCAGTG 1260  
QY 1261 GATGGTCACCTGTCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGGCTGTGGTTC 1320  
DB 1261 GATGGTCACCTGTCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGGCTGTGGTTC 1320  
QY 1321 TTTCCCTCTGGAGTGTGCTCCACATCTCTGGCTGGGTGTACAGAAACAGGAAAGGG 1380  
DB 1321 TTTCCCTCTGGAGTGTGCTCCACATCTCTGGCTGGGTGTACAGAAACAGGAAAGGG 1380  
QY 1381 TCCCGCAGCTCTGGATGCCAAAGTCAAGCTGATTTGCCAACACTTTGTGCAACTCCCGC 1440  
DB 1381 TCCCGCAGCTCTGGATGCCAAAGTCAAGCTGATTTGCCAACACTTTGTGCAACTCCCGC 1440  
QY 1441 CAATCTATGACCAATGATTTGATGACAGTATGATGTCGAGGAATCTTCAGAAACCT 1500  
DB 1441 CAATCTATGACCAATGATTTGATGACAGTATGATGTCGAGGAATCTTCAGAAACCT 1500  
QY 1501 GGGCAAGACCTCCAGGGTGTCTGGAGGCCCCCTGACCTGTGAGAGAGAGCCGACCC 1560  
DB 1501 GGGCAAGACCTCCAGGGTGTCTGGAGGCCCCCTGACCTGTGAGAGAGAGCCGACCC 1560  
QY 1561 TACTACGTCTATGGATAGTGTGGGCTGTGGAGTGTGGAGAGAGCCAGGGTCTAC 1620  
DB 1561 TACTACGTCTATGGATAGTGTGGGCTGTGGAGTGTGGAGAGAGCCAGGGTCTAC 1620  
QY 1621 ACCCAAGTACCAATCTCTGAATTTGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680  
DB 1621 ACCCAAGTACCAATCTCTGAATTTGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680  
QY 1681 TAA 1683  
DB 1681 TAA 1683

RESULT 2  
US-10-391-215-1  
Sequence 1, Application US/10391215  
Publication No. US20040009543A1  
GENERAL INFORMATION:  
APPLICANT: KIEHL, STEFAN  
APPLICANT: WILLEIT, JOHANN  
APPLICANT: WIEDERMANN, CHRISTIAN JOSEF  
APPLICANT: ROEMISCH, JUERGEN  
APPLICANT: WEIMER, THOMAS  
APPLICANT: FEUSSNER, ANNETTE  
APPLICANT: STOEHR, HANS-ARNOLD  
APPLICANT: DOERSAM, VOLKER  
TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE  
FILE REFERENCE: 06478.1457-01  
CURRENT FILING DATE: 2003-03-19  
CURRENT FILING DATE: 2003-03-19  
PRIOR FILING DATE: 2001-07-26  
PRIOR FILING DATE: 2001-07-26  
PRIOR FILING DATE: 2000-07-26  
PRIOR FILING DATE: 2000-07-26  
PRIOR FILING DATE: 2000-10-10  
PRIOR FILING DATE: 2000-10-10  
PRIOR FILING DATE: 2000-10-21  
PRIOR FILING DATE: 2000-10-21  
PRIOR FILING DATE: 2001-04-12

QY 1 ATGTTTCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGTTGGAAAGACA 60  
DB 1 ATGTTTCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGTTGGAAAGACA 60  
QY 61 GCTGTGGGTTCTCCCTGATGCTTTATTTGAAAGCCCTGGACCCAGACTGACCCCTGAC 120  
DB 61 GCTGTGGGTTCTCCCTGATGCTTTATTTGAAAGCCCTGGACCCAGACTGACCCCTGAC 120  
QY 121 CAGTATGATTACAGCTTACAGGATTAATATCAGGAAGAGAAACACAGTAGACACTTACC 180  
DB 121 CAGTATGATTACAGCTTACAGGATTAATATCAGGAAGAGAAACACAGTAGACACTTACC 180  
QY 181 CATGCTGAGAAATCCTGACTGCTACTACCTGAGGACCAAGCTGATCCATGCGCCCAAC 240  
DB 181 CATGCTGAGAAATCCTGACTGCTACTACCTGAGGACCAAGCTGATCCATGCGCCCAAC 240  
QY 241 CCCTGTGAAACAGGTTGGGACTGCTCTGCTGAGGACCACTTACATGCAAGCTGCTG 300  
DB 241 CCCTGTGAAACAGGTTGGGACTGCTCTGCTGAGGACCACTTACATGCAAGCTGCTG 300  
QY 301 GCTCCTTTCTCTGGGAATAGTGTGCAAGAGTGCAGAAATACGTCGAGGACCAACCATGT 360  
DB 301 GCTCCTTTCTCTGGGAATAGTGTGCAAGAGTGCAGAAATACGTCGAGGACCAACCATGT 360  
QY 361 GCGGGGGCAATGCTCTATCCAGAGTCCCTCTACTACCGCTGTGTCTGTAAACAC 420  
DB 361 GCGGGGGCAATGCTCTATCCAGAGTCCCTCTACTACCGCTGTGTCTGTAAACAC 420  
QY 421 CCTTACACAGTCTCCAGCTCTCCCAAGTGGTTCCTGTATGAGGCGAAACCCCTGCCAG 480  
DB 421 CCTTACACAGTCTCCAGCTCTCCCAAGTGGTTCCTGTATGAGGCGAAACCCCTGCCAG 480  
QY 481 AATGGGCTACTGCTCCCGCATAGCGAGATCCAGTTCACCTGCTGCTGCTCCGAC 540  
DB 481 AATGGGCTACTGCTCCCGCATAGCGAGATCCAGTTCACCTGCTGCTGCTCCGAC 540  
QY 541 CAGTTCAAGGGAAATCTGTGAAATAGTTCCTGATGCTGCTATGTTGGCGATGCTAC 600  
DB 541 CAGTTCAAGGGAAATCTGTGAAATAGTTCCTGATGCTGCTATGTTGGCGATGCTAC 600  
QY 601 TCTTACGAGGAAATGATAGACAGTCAACAGCATGCTGCTTTACTGGAATCC 660  
DB 601 TCTTACGAGGAAATGATAGACAGTCAACAGCATGCTGCTTTACTGGAATCC 660  
QY 661 CACTCTCTTTCAGGAAATTAACAATGTTATGAGGATGCTGAAACCCATGGGATT 720  
DB 661 CACTCTCTTTCAGGAAATTAACAATGTTATGAGGATGCTGAAACCCATGGGATT 720  
QY 721 GGGAAACAATTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGCTTTATTA 780  
DB 721 GGGAAACAATTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGCTTTATTA 780  
QY 781 GTTACCAATGACAGTGAATGGGAATCTGTGATGCTCAGGCTGCTCAGCCAGGAC 840  
DB 781 GTTACCAATGACAGTGAATGGGAATCTGTGATGCTCAGGCTGCTCAGCCAGGAC 840  
QY 841 GTTGCTTACCAGGAAAGCCCACTGAGCAATCAACCAAGTTCGGGGTTGACTCC 900  
DB 841 GTTGCTTACCAGGAAAGCCCACTGAGCAATCAACCAAGTTCGGGGTTGACTCC 900  
QY 901 TGTGAAAGACTGATAGCAGAGAGAGATCAAGAAATCTATGAGGCTTTAAGAGC 960  
DB 901 TGTGAAAGACTGATAGCAGAGAGAGATCAAGAAATCTATGAGGCTTTAAGAGC 960  
QY 961 ACGGGGGCAAGCAACCCATGCGAGGCTCCCTCCAGTCTCTGCTGCTCAGCAATCTCC 1020  
DB 961 ACGGGGGCAAGCAACCCATGCGAGGCTCCCTCCAGTCTCTGCTGCTCAGCAATCTCC 1020  
QY 1021 ATGCCCCAGGCACTTCTGTGGTGGGCTGATCCACCCCTGCTGGGCTCAGCTGCT 1080  
DB 1021 ATGCCCCAGGCACTTCTGTGGTGGGCTGATCCACCCCTGCTGGGCTCAGCTGCT 1080

; NUMBER OF SEQ ID NOS: 8									
; SOFTWARE: PatentIn Ver. 2.1									
; SEQ ID NO 1									
; LENGTH: 1683									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
US-10-391-215-1									
Query Match 100.0%; Score 1683; DB 16; Length 1683;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	ATGTTTCCAGAGTCTGATCTCATGTTCTGCTGTTAATGGCTCTGTTGGGAAGACA	60						
Db	1	ATGTTTCCAGAGTCTGATCTCATGTTCTGCTGTTAATGGCTCTGTTGGGAAGACA	60						
Qy	61	GCCTGTGGTTCCTCCCTGATGCTCTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC	120						
Db	61	GCCTGTGGTTCCTCCCTGATGCTCTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC	120						
Qy	121	CAGTATGATTACAGCTACAGAGTATTAAATCAGGAAGAGACACAGTAGGACACTTACC	180						
Db	121	CAGTATGATTACAGCTACAGAGTATTAAATCAGGAAGAGACACAGTAGGACACTTACC	180						
Qy	181	CATGCTGAGAAATCCTGACTGTGTTACTACACTGAGGACCAAGCTGATCCATGCGACCCCAAC	240						
Db	181	CATGCTGAGAAATCCTGACTGTGTTACTACACTGAGGACCAAGCTGATCCATGCGACCCCAAC	240						
Qy	241	CCCTGTGAACACGGTGGGAGTGCCTCTGTCATGGAGCACCTTCACATGACAGTGCCTG	300						
Db	241	CCCTGTGAACACGGTGGGAGTGCCTCTGTCATGGAGCACCTTCACATGACAGTGCCTG	300						
Qy	301	GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAAAATAGCTGCAAGGACCAACCCATGT	360						
Db	301	GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAAAATAGCTGCAAGGACCAACCCATGT	360						
Qy	361	GGCCGGGCGCAATGTCCTATTACCCAGAGTCTCCCTACTACCCGCTGTGTGTTAAACAC	420						
Db	361	GGCCGGGCGCAATGTCCTATTACCCAGAGTCTCCCTACTACCCGCTGTGTGTTAAACAC	420						
Qy	421	CCTTACACAGTCCAGTGTCTCCAGTGTCTGTATGACGCGCAAAACCCCTGCCAG	480						
Db	421	CCTTACACAGTCCAGTGTCTCCAGTGTCTGTATGACGCGCAAAACCCCTGCCAG	480						
Qy	481	AATGGGGTACCTGCTCCCGGCATAAAGCGGAGATCCAAAGTTCACTGTGCTGTCGCCGAC	540						
Db	481	AATGGGGTACCTGCTCCCGGCATAAAGCGGAGATCCAAAGTTCACTGTGCTGTCGCCGAC	540						
Qy	541	CAGTTCAAGGGGAATTCCTGCAATAGTGTCTGATGACTGCTATGTTGGCGATGGCTAC	600						
Db	541	CAGTTCAAGGGGAATTCCTGCAATAGTGTCTGATGACTGCTATGTTGGCGATGGCTAC	600						
Qy	601	TCTTACCAGGGGAAATGAATAGGACAGTCAACAGCATGCGTGCCTTTACTGGAACTCC	660						
Db	601	TCTTACCAGGGGAAATGAATAGGACAGTCAACAGCATGCGTGCCTTTACTGGAACTCC	660						
Qy	661	CACCTCCTCTTCAGGAGAAATTAACAATGTTTATGGAGGATGCTGAAACCCATGGGATT	720						
Db	661	CACCTCCTCTTCAGGAGAAATTAACAATGTTTATGGAGGATGCTGAAACCCATGGGATT	720						
Qy	721	GGGGAACCAATTCCTGCAGAAACCCAGATGCGGACGAAAGCCCTGCTGTTATTAAA	780						
Db	721	GGGGAACCAATTCCTGCAGAAACCCAGATGCGGACGAAAGCCCTGCTGTTATTAAA	780						
Qy	781	GTACCAATGACAAGGTGAATGGGAATGCTGTGATGTTCTCAGCTGCTCAGCCAGGAC	840						
Db	781	GTACCAATGACAAGGTGAATGGGAATGCTGTGATGTTCTCAGCTGCTCAGCCAGGAC	840						
Qy	841	GTTCCTACCCAGAGGAAGCCCTCAGCCATCAACCAAGCTTCCGGGTTTGACTCC	900						
Db	841	GTTCCTACCCAGAGGAAGCCCTCAGCCATCAACCAAGCTTCCGGGTTTGACTCC	900						
Qy	901	TGTGAAAGACTGATAGACAGAGGAGATCAAGAGAATCTATGGAGGCTTTAAGAGC	960						

RESULT 3  
US-09-880-107-1668  
; Sequence 1668, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherif, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02

; NUMBER OF SEQ ID NOS: 3950			
; SOFTWARE: PatertIn Ver. 2.1			
; SEQ ID NO 1668			
; LENGTH: 3008			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D49742			
US-09-880-107-1668			
Query Match 100.0%; Score 1683; DB 9; Length 3008;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	ATGTTGGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGTTGGGAAAGACA	60
DB	97	ATGTTGGCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGTTGGGAAAGACA	156
QY	61	GCTGTGGGTTCTCCCTGATGCTTTTATTTGGAAGCTGGACCCAGACTGGACCCCTGAC	120
DB	157	GCTGTGGGTTCTCCCTGATGCTTTTATTTGGAAGCTGGACCCAGACTGGACCCCTGAC	216
QY	121	CAGTATGATTACAGCTACGAGTACGAGGATTATATCAGGAAGAGAAACACAGTAGACACTTACC	180
DB	217	CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAGAAACACAGTAGACACTTACC	276
QY	181	CATGCTGAGATCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC	240
DB	277	CATGCTGAGATCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC	336
QY	241	CCCTGTGAACACGGTGGGAGCTCCCTGCTCCATGGGAGCACCTTACATGACAGTGCCTG	300
DB	337	CCCTGTGAACACGGTGGGAGCTCCCTGCTCCATGGGAGCACCTTACATGACAGTGCCTG	396
QY	301	GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAAAATACGTGCAAGGACCAACCATGT	360
DB	397	GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAAAATACGTGCAAGGACCAACCATGT	456
QY	361	GGCGGGGGCAATGCTCTATTACCCAGAGTCTCTCCCTACTACCGCTGTCTGTAAACAC	420
DB	457	GGCGGGGGCAATGCTCTATTACCCAGAGTCTCTCCCTACTACCGCTGTCTGTAAACAC	516
QY	421	CCTTACACAGTCCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGCAAAACCCCTGCCAG	480
DB	517	CCTTACACAGTCCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGCAAAACCCCTGCCAG	576
QY	481	AATGGGGTACCTGCTCCGGCATAGCGAGATCCAGTTCACCTGTCCTGTCCTCCGAC	540
DB	577	AATGGGGTACCTGCTCCGGCATAGCGAGATCCAGTTCACCTGTCCTGTCCTCCGAC	636
QY	541	CAGTTCAAGGGGAAATTTCTGAAATAGGTTCTGATGACTGCTATGTTGGCGATGGCTAC	600
DB	637	CAGTTCAAGGGGAAATTTCTGAAATAGGTTCTGATGACTGCTATGTTGGCGATGGCTAC	696
QY	601	TCTTACCGGGAATGATAGACAGTCAACAGCATGCGTGCCTTTACTGGACTCC	660
DB	697	TCTTACCGGGAATGATAGACAGTCAACAGCATGCGTGCCTTTACTGGACTCC	756
QY	661	CACCTCCTCTTGCAGGGAATTAACAATGTTTATGGAGGATGCTGAAACCCCATGGATT	720
DB	757	CACCTCCTCTTGCAGGGAATTAACAATGTTTATGGAGGATGCTGAAACCCCATGGATT	816
QY	721	GGGGAACACAAATTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGGTCTTTATAA	780
DB	817	GGGGAACACAAATTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGGTCTTTATAA	876
QY	781	GTTACCAATGACAAGGTGAAATGGGAATACGTGTGATGCTCAGCCTGCTCAGCCAGGAC	840
DB	877	GTTACCAATGACAAGGTGAAATGGGAATACGTGTGATGCTCAGCCTGCTCAGCCAGGAC	936
QY	841	GTTGCCCTACCCAGAGGAAGCCCACTGAGCCATCAACAGCTTCCGGGGTTGACTCC	900
DB	937	GTTGCCCTACCCAGAGGAAGCCCACTGAGCCATCAACAGCTTCCGGGGTTGACTCC	996

QY	901	TGTGGAAGACTGAGATAGCAGAGAGGAAGATCAAGAGAATCTATGAGGCTTTAAGAGC	960
DB	997	TGTGGAAGACTGAGATAGCAGAGAGGAAGATCAAGAGAATCTATGAGGCTTTAAGAGC	1056
QY	961	ACGGCGGGCAAGCAACCATGGCAGGCTGCTCCAGTCTCGCTGCTGCTGACATCTCC	1020
DB	1057	ACGGCGGGCAAGCAACCATGGCAGGCTGCTCCAGTCTCGCTGCTGCTGACATCTCC	1116
QY	1021	ATGCCCCAGGCGCACCTTCTGTGTGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT	1080
DB	1117	ATGCCCCAGGCGCACCTTCTGTGTGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT	1176
QY	1081	GCCCACTGCAACGACATAAATAACAGACATCTTAAAGTGGTGTAGGGACCAAGGACCTG	1140
DB	1177	GCCCACTGCAACGACATAAATAACAGACATCTTAAAGTGGTGTAGGGACCAAGGACCTG	1236
QY	1141	AAGAAGACGAATTTCTATGAGCAGAGCTTTAGGAGTGGTGTAGGGACCAAGGACCTG	1200
DB	1237	AAGAAGACGAATTTCTATGAGCAGAGCTTTAGGAGTGGTGTAGGGACCAAGGACCTG	1296
QY	1201	TACAATGAAGAGATGAGATTCCCAATATGATATTCATGCTCAAGTTAAAGCCAGTG	1260
DB	1297	TACAATGAAGAGATGAGATTCCCAATATGATATTCATGCTCAAGTTAAAGCCAGTG	1356
QY	1261	GATGCTCACTGCTCTAGATCCAAATACGTGAAGACTGTGTGCTGCTGATGGTCC	1320
DB	1357	GATGCTCACTGCTCTAGATCCAAATACGTGAAGACTGTGTGCTGCTGATGGTCC	1416
QY	1321	TTTCCCTCTGGGAGTGAAGTCCACATCTCTGGCTGGGGTGTACAGAAAACAGAAAAGGG	1380
DB	1417	TTTCCCTCTGGGAGTGAAGTCCACATCTCTGGCTGGGGTGTACAGAAAACAGAAAAGGG	1476
QY	1381	TCGGCCAGCTCTGAGTCCAAAGTCAAGTATGTCGCAACACTTTGCAACTCCCGC	1440
DB	1477	TCGGCCAGCTCTGAGTCCAAAGTCAAGTATGTCGCAACACTTTGCAACTCCCGC	1536
QY	1441	CAACTCTATGACACATGATTGATGACAGTATGATCTGTGCAAGAAATCTTCAGAAAACCT	1500
DB	1537	CAACTCTATGACACATGATTGATGACAGTATGATCTGTGCAAGAAATCTTCAGAAAACCT	1596
QY	1501	GGGCAAGACACTGCCAGGCTGACTCTGGAGGCCCCCTGACCTGTGAGAGGACGGACC	1560
DB	1597	GGGCAAGACACTGCCAGGCTGACTCTGGAGGCCCCCTGACCTGTGAGAGGACGGACC	1656
QY	1561	TACTACGCTCTATGGATAGTGAAGTGGGCTGGAGTGGGAGAGGCGGAGGCTTAC	1620
DB	1657	TACTACGCTCTATGGATAGTGAAGTGGGCTGGAGTGGGAGAGGCGGAGGCTTAC	1716
QY	1621	ACCAAGTTACCAAAATTCCTGAATTTGGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC	1680
DB	1717	ACCAAGTTACCAAAATTCCTGAATTTGGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC	1776
QY	1681	TAA	1681
DB	1777	TAA	1779

RESULT 4  
US-10-391-215-2  
; Sequence 2. Application US/10391215  
; Publication No. US2004009543A1  
; GENERAL INFORMATION:  
; APPLICANT: KIEHL, STEFAN  
; APPLICANT: WILLEIT, JOHANN  
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF  
; APPLICANT: ROEMISCH, JUERGEN  
; APPLICANT: WEIMER, THOMAS  
; APPLICANT: FEUSSNER, ANNETTE  
; APPLICANT: STOEHR, HANS-ARNOLD  
; APPLICANT: DOERSAM, VOLKER  
; TITLE OF INVENTION: (FSAP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS

FILE REFERENCE: 06478.1457-01									
CURRENT APPLICATION NUMBER: US/10/391.215									
CURRENT FILING DATE: 2003-03-19									
PRIORITY FILING DATE: 2003-03-19									
PRIORITY FILING DATE: 2001-07-26									
PRIORITY FILING DATE: 2001-07-26									
PRIORITY FILING DATE: 2000-07-26									
PRIORITY FILING DATE: 2000-07-26									
PRIORITY FILING DATE: 2000-10-10									
PRIORITY FILING DATE: 2000-10-10									
PRIORITY FILING DATE: 2000-10-21									
PRIORITY FILING DATE: 2001-04-12									
NUMBER OF SEQ ID NOS: 8									
SOFTWARE: PatentIn Ver. 2.1									
SEQ ID NO 2									
LENGTH: 1683									
TYPE: DNA									
ORGANISM: Homo sapiens									
US-10-391-215-2									
Query Match 99.9%; Score 1681.4; DB 16; Length 1683;									
Best Local Similarity 99.9%; Pred. No. 0;									
Matches 1682; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
QY	1	ATGTTTGCAGGATCTCTGATCTCCATGTTCTCTGTTTAAATGGCTCTGGTGGGAAAGACA	60						
DB	1	ATGTTTGCAGGATCTCTGATCTCCATGTTCTGTTTAAATGGCTCTGGTGGGAAAGACA	60						
QY	61	GCCTGTGGTTCCTCCCTGATGCTTTATTGAAAGCCTGGACCCAGAGCTGGACCCCTGAC	120						
DB	61	GCCTGTGGTTCCTCCCTGATGCTTTATTGAAAGCCTGGACCCAGAGCTGGACCCCTGAC	120						
QY	121	CAGTATGATTACAGCTACGAGGATTAATCAGGAAGAGAACACACAGTAGACACTTACC	180						
DB	121	CAGTATGATTACAGCTACGAGGATTAATCAGGAAGAGAACACACAGTAGACACTTACC	180						
QY	181	CATGCTGGAATCTCTGATGCTTACACTGAGGACCAAGCTGATCCATGAGCTGCTG	240						
DB	181	CATGCTGGAATCTCTGATGCTTACACTGAGGACCAAGCTGATCCATGAGCTGCTG	240						
QY	241	CCCTGTGAACACGGTGGGAGCTGCTCTGATGAGGACACCTTCCATGAGCTGCTG	300						
DB	241	CCCTGTGAACACGGTGGGAGCTGCTCTGATGAGGACACCTTCCATGAGCTGCTG	300						
QY	301	GCTCCTTTCTGGAATTAAGTGTGAGAAAGTGCAAAATACGTGCAAGGACACCCATGT	360						
DB	301	GCTCCTTTCTGGAATTAAGTGTGAGAAAGTGCAAAATACGTGCAAGGACACCCATGT	360						
QY	361	GGCCGGGGCCATGTCTCATTACCCAGAGTCTCCCTACTACCGCTGTGTCTGTAACAC	420						
DB	361	GGCCGGGGCCATGTCTCATTACCCAGAGTCTCCCTACTACCGCTGTGTCTGTAACAC	420						
QY	421	CCTTACAGGTCCTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACCTGTGCTGCCGAC	480						
DB	421	CCTTACAGGTCCTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACCTGTGCTGCCGAC	480						
QY	481	AATGGGGTCTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACCTGTGCTGCCGAC	540						
DB	481	AATGGGGTCTGCTCCCGGATTAAGCGGAGATCCAAAGTTCACCTGTGCTGCCGAC	540						
QY	541	CAGTTCAAGGGAAATCTGTGAATAGTTCGATGAGCTGTGTTGGGATGGCTTAC	600						
DB	541	CAGTTCAAGGGAAATCTGTGAATAGTTCGATGAGCTGTGTTGGGATGGCTTAC	600						
QY	601	TCTTACCGAGGAAATGAATAGGACAGTCAACAGCAGCTGCTTACTTGGAACTCC	660						
DB	601	TCTTACCGAGGAAATGAATAGGACAGTCAACAGCAGCTGCTTACTTGGAACTCC	660						
QY	661	CACCTCCTCTTCAGGAGAAATACACATGTTTATGGAGATGCTGAAACCCATGGGATT	720						
DB	661	CACCTCCTCTTCAGGAGAAATACACATGTTTATGGAGATGCTGAAACCCATGGGATT	720						

RESULT 5  
US-10-391-215-3  
; Sequence 3, Application US/10391215  
; Publication No. US20040009543A1

QY	721	GGGGAACACAAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGCTGCTTTTAAA	780
DB	721	GGGGAACACAAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGCTGCTTTTAAA	780
QY	781	GTTACCAATGACAAAGTGAATGGAAATCTGTGATGCTCTAGCTGCTCAGCCCGAGAC	840
DB	781	GTTACCAATGACAAAGTGAATGGAAATCTGTGATGCTCTAGCTGCTCAGCCCGAGAC	840
QY	841	GTTGCTTACCCAGAGGAAACCCCACTAGCCATCAACCAAGCTTCCGGGGTTTGAATCC	900
DB	841	GTTGCTTACCCAGAGGAAACCCCACTAGCCATCAACCAAGCTTCCGGGGTTTGAATCC	900
QY	901	TGTGAAAGACTGAGATAGCAGAGAGAGATCAAGAGATCTATGAGGCTTTAAGAGC	960
DB	901	TGTGAAAGACTGAGATAGCAGAGAGAGATCAAGAGATCTATGAGGCTTTAAGAGC	960
QY	961	ACGGCGGCAAGCACCCATCGGAGGCTCCCTCCAGTCTCTGCTGCTCTGACCATCTCC	1020
DB	961	ACGGCGGCAAGCACCCATCGGAGGCTCCCTCCAGTCTCTGCTGCTCTGACCATCTCC	1020
QY	1021	ATGCCCCAGGGGCACTTCTGTGTGGGGCTGTGATCCACCCCTGCTGCTGCTGCTGCT	1080
DB	1021	ATGCCCCAGGGGCACTTCTGTGTGGGGCTGTGATCCACCCCTGCTGCTGCTGCTGCT	1080
QY	1081	GCCCACTGACCGACATTAACCAACACATCTAAAGTGTGCTAGGGACCCAGGACCTG	1140
DB	1081	GCCCACTGACCGACATTAACCAACACATCTAAAGTGTGCTAGGGACCCAGGACCTG	1140
QY	1141	AAGAAAGAGAAATTTTATGAGCAGAGCTTTAGGTTGGAGAGATATTCAGGTACAGCCAC	1200
DB	1141	AAGAAAGAGAAATTTTATGAGCAGAGCTTTAGGTTGGAGAGATATTCAGGTACAGCCAC	1200
QY	1201	TACATGAAGAGATGAGATTTCCCAATGATATTCGATTCCTCAAGTTAAGCCAGTG	1260
DB	1201	TACATGAAGAGATGAGATTTCCCAATGATATTCGATTCCTCAAGTTAAGCCAGTG	1260
QY	1261	GATGCTCAGTGTCTCTAGAAATCAAGTGAAGACTGTGTGCTTGCTGCTGCTGCTGCT	1320
DB	1261	GATGCTCAGTGTCTCTAGAAATCAAGTGAAGACTGTGTGCTTGCTGCTGCTGCTGCT	1320
QY	1321	TTTCCCTCTGGGAGTGAAGTGCACATCTCTGGTGGGGTGTACAGAAACAGAAAGGG	1380
DB	1321	TTTCCCTCTGGGAGTGAAGTGCACATCTCTGGTGGGGTGTACAGAAACAGAAAGGG	1380
QY	1381	TCCCGCAGCTCTCTGATGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	1440
DB	1381	TCCCGCAGCTCTCTGATGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	1440
QY	1441	CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1500
DB	1441	CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1500
QY	1501	GGGCAAGACCTGCGAGGCTGCTGAGGCGCCCTGACCTGTGAGAGGACCGCAC	1560
DB	1501	GGGCAAGACCTGCGAGGCTGCTGAGGCGCCCTGACCTGTGAGAGGACCGCAC	1560
QY	1561	TACTACGCTTATGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1620
DB	1561	TACTACGCTTATGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1620
QY	1621	ACCAAGTTACCAATTTCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC	1680
DB	1621	ACCAAGTTACCAATTTCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC	1680
QY	1681	TAA 1683	
DB	1681	TAA 1683	

GENERAL INFORMATION:	
APPLICANT: KIECHL, STEFAN	
APPLICANT: WILLEIT, JOHANN	
APPLICANT: WIEDERMANN, CHRISTIAN JOSEF	
APPLICANT: ROEMISCH, JUERGEN	
APPLICANT: WEIMER, THOMAS	
APPLICANT: FEUSSNER, ANNETTE	
APPLICANT: STOEHR, HANS-ARNOLD	
APPLICANT: DOERSAM, VOLKER	
TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEINASE	
FILE OF INVENTION: (PSAP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS	
FILE REFERENCE: 06478.1457-01	
CURRENT APPLICATION NUMBER: US/10/391,215	
CURRENT FILING DATE: 2003-03-19	
PRIORITY APPLICATION NUMBER: 09/912,559	
PRIORITY FILING DATE: 2001-07-26	
PRIORITY FILING DATE: 2000-07-26	
PRIORITY FILING DATE: 2000-10-10	
PRIORITY FILING DATE: 2000-10-21	
PRIORITY FILING DATE: 2001-04-12	
NUMBER OF SEQ ID NOS: 8	
SOFTWARE: PatentIn Ver. 2.1	
SEQ ID NO 3	
LENGTH: 1683	
TYPE: DNA	
ORGANISM: Homo sapiens	
US-10-391-215-3	
Query Match 99.9%; Score 1681.4; DB 16; Length 1683;	
Best Local Similarity 99.9%; Pred. No. 0;	
Matches 1682; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 ATGTTTCCAGGATGCTGATCTCCATGTTCTGTTGTTAATGGCTCTGGTGGGAAGACA 60
DB	1 ATGTTTCCAGGATGCTGATCTCCATGTTCTGTTAATGGCTCTGGTGGGAAGACA 60
QY	61 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCCTGGACCCAGCTGGACCCCTGAC 120
DB	61 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAGCCTGGACCCAGCTGGACCCCTGAC 120
QY	121 CAGTATGATTACAGCTACGAGGATTATAATCAGGAAGAACCAAGTACGACACTTACC 180
DB	121 CAGTATGATTACAGCTACGAGGATTATAATCAGGAAGAACCAAGTACGACACTTACC 180
QY	181 CATGCTGAGAACTCTGACTGGTACTACATGAGGACCAAGTGCATGATCCAGCCCAAC 240
DB	181 CATGCTGAGAACTCTGACTGGTACTACATGAGGACCAAGTGCATGATCCAGCCCAAC 240
QY	241 CCCTGTGAAACACGGTGGGACTGCTGCTCCATGGAGACCTTTCATGACGCTGCTG 300
DB	241 CCCTGTGAAACACGGTGGGACTGCTGCTCCATGGAGACCTTTCATGACGCTGCTG 300
QY	301 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTCAAAATAGTGCAGGCAAAACCCATGT 360
DB	301 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTCAAAATAGTGCAGGCAAAACCCATGT 360
QY	361 GGCGGGGCAATGCTCTATACCCAGAGTCCCTCCCTACTACGCTGCTGTCTGTAAACAC 420
DB	361 GGCGGGGCAATGCTCTATACCCAGAGTCCCTCCCTACTACGCTGCTGTCTGTAAACAC 420
QY	421 CTTTACACAGGTCCAGCTGCTCCCAAGTGGTTCCTGTATGACGGCCAAACCCCTGCCAG 480
DB	421 CTTTACACAGGTCCAGCTGCTCCCAAGTGGTTCCTGTATGACGGCCAAACCCCTGCCAG 480
QY	481 AATGGGGTACTGCTCCGGCATAGCGGAGATCCAAAGTTCACCTGTGCTGTCCCGAC 540
DB	481 AATGGGGTACTGCTCCGGCATAGCGGAGATCCAAAGTTCACCTGTGCTGTCCCGAC 540
QY	541 CAGTTCAAGGGGAAATTTCTGTGAAATAGGTTCTGATGACTGCTATGTTGGCGATGCTTAC 600

541	CAGTTCAAGGGGAAATTTCTGTGAAATAGGTTCTGATGACTGCTATGTTGGCGATGGCTAC	600
601	TCATTACCGAGGGAATATGATAGGACAGTCAACACGATCGCTGCTTACTTGGAACTCC	660
601	TCATTACCGAGGGAATATGATAGGACAGTCAACACGATCGCTGCTTACTTGGAACTCC	660
661	CACCTCCTCTTGAGGAGAAATTAACAATGTTTATGAGAGTCTGCTGAAACCCATGGGATT	720
661	CACCTCCTCTTGAGGAGAAATTAACAATGTTTATGAGAGTCTGCTGAAACCCATGGGATT	720
721	GGGGAACACAATTTCTGCAGAAACCCAGATGCGGAGCAAAAGCCCTGGTGTCTTTATAA	780
721	GGGGAACACAATTTCTGCAGAAACCCAGATGCGGAGCAAAAGCCCTGGTGTCTTTATAA	780
781	GTTACCAATGACAAGGTGAAATGGGAATACTGTGATGTTCTCAGCTGCTCAGCCAGGAC	840
781	GTTACCAATGACAAGGTGAAATGGGAATACTGTGATGTTCTCAGCTGCTCAGCCAGGAC	840
841	GTTGCTTACCAGAGGAAGCCCACTGAGCCATCAACCAAGCTTCCGGGTTTGACTCC	900
841	GTTGCTTACCAGAGGAAGCCCACTGAGCCATCAACCAAGCTTCCGGGTTTGACTCC	900
901	TGTGGAAGAGCTGAGATAGCAGAGAGGAAGATCAAGAGAAATCTATGAGGCTTTAAGAC	960
901	TGTGGAAGAGCTGAGATAGCAGAGAGGAAGATCAAGAGAAATCTATGAGGCTTTAAGAC	960
961	ACGGCGGCAAGCAACCATGCGAGGCTCCCTCCAGTCCCTGCTGCTGCTGCTGCTGCTGCT	1020
961	ACGGCGGCAAGCAACCATGCGAGGCTCCCTCCAGTCCCTGCTGCTGCTGCTGCTGCTGCT	1020
1021	ATGCCCCAGGCGCACTTCTGTGTGGGGCGGTGATCCACCCCTGCTGGGTGCTCACTGCT	1080
1021	ATGCCCCAGGCGCACTTCTGTGTGGGGCGGTGATCCACCCCTGCTGGGTGCTCACTGCT	1080
1081	GCCCACTGCAACGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT	1140
1081	GCCCACTGCAACGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT	1140
1141	AAGAAAGAGAAATTTCTGAGCAGAGCTTTAGGCTGGAGAGATATTTCAAGTACAGCCAC	1200
1141	AAGAAAGAGAAATTTCTGAGCAGAGCTTTAGGCTGGAGAGATATTTCAAGTACAGCCAC	1200
1201	TACAAATGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT	1260
1201	TACAAATGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT	1260
1261	GATGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1320
1261	GATGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1320
1321	TTTCCCTCTGGAGTGAAGTGCACATCTCTGGCTGGGCTGTTACAGAAACAGGAAAGGG	1380
1321	TTTCCCTCTGGAGTGAAGTGCACATCTCTGGCTGGGCTGTTACAGAAACAGGAAAGGG	1380
1381	TCGGCGGCAAGTCTGAGTCCAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT	1440
1381	TCGGCGGCAAGTCTGAGTCCAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT	1440
1441	CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1500
1441	CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1500
1501	GGGGAAGACACTGCGAGGAGTCTGAGGCGCCCTGACCTGTGAGAGAGAGAGAGAGAGAG	1560
1501	GGGGAAGACACTGCGAGGAGTCTGAGGCGCCCTGACCTGTGAGAGAGAGAGAGAGAGAG	1560
1561	TACTACGCTATGAGGATGAGTGGGCTGGAGTGTGGGAAAGAGAGAGAGAGAGAGAGTAC	1620
1561	TACTACGCTATGAGGATGAGTGGGCTGGAGTGTGGGAAAGAGAGAGAGAGAGAGAGTAC	1620
1621	ACCAAAGTTACAAATTTCTGTAATGATCAAGAGCCCAATCAAAAGTGAAGTGGCTTC	1680



Db 1621 ACCCAGTTACCAATTCCTGATCGATCAAGCCACCATCAAAAGTGAAGTGCTTC 1680  
 QY 1681 TAA 1683  
 Db 1681 TAA 1683

RESULT 6

US-09-912-559-2  
 ; Sequence 2, Application US/09912559  
 ; Patent No. US20020142316A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ROEMISCH, JUERGEN  
 ; APPLICANT: STOEHR, HANS-ARNOLD  
 ; APPLICANT: FEUSSNER, ANNETTE  
 ; APPLICANT: LANG, WIEGAND  
 ; APPLICANT: WEIMER, THOMAS  
 ; APPLICANT: BECKER, MARGRET  
 ; APPLICANT: NERLICH, CLAUDIA  
 ; APPLICANT: MUTH-NAUMANN, GUDRUN  
 ; TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND  
 ; TITLE OF INVENTION: DETECTION METHODS USING SPECIFIC ANTIBODIES  
 ; FILE REFERENCE: 06478.1457  
 ; CURRENT APPLICATION NUMBER: US/09/912,559  
 ; CURRENT FILING DATE: 2001-07-26  
 ; PRIOR APPLICATION NUMBER: DE 100 36 641.4  
 ; PRIOR FILING DATE: 2000-07-26  
 ; PRIOR APPLICATION NUMBER: DE 100 50 040.4  
 ; PRIOR FILING DATE: 2000-10-10  
 ; PRIOR APPLICATION NUMBER: DE 100 52 319.6  
 ; PRIOR FILING DATE: 2000-10-21  
 ; PRIOR APPLICATION NUMBER: DE 101 18 706.8  
 ; PRIOR FILING DATE: 2001-04-12  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 1683  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-912-559-2

Query Match 99.8%; Score 1679.8; DB 9; Length 1683;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 ATGTTGGCAGGATGCTGATCCATGTTCTGCTGTTAATGGCTCTGTTGGGAAGACA 60  
 Db 1 ATGTTGGCAGGATGCTGATCCATGTTCTGCTGTTAATGGCTCTGTTGGGAAGACA 60  
 QY 61 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAAGCTGGACCCAGACTGGACCCCTGAC 120  
 Db 61 GCCTGTGGGTTCTCCCTGATGCTTTATTGGAAAGCTGGACCCAGACTGGACCCCTGAC 120  
 QY 121 CAGTATGATTACAGCTACGAGGATTAATACAGGAAGACAACACAGTAGACACTTACC 180  
 Db 121 CAGTATGATTACAGCTACGAGGATTAATACAGGAAGACAACACAGTAGACACTTACC 180  
 QY 181 CATGCTGAGAACTCTGACTGCTACTACCTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240  
 Db 181 CATGCTGAGAACTCTGACTGCTACTACCTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240  
 QY 241 CCCTGTGAACACCGTGGGACTCCCTGCTCCATGAGGACACTTCAATGACGAGTCCCTG 300  
 Db 241 CCCTGTGAACACCGTGGGACTCCCTGCTCCATGAGGACACTTCAATGACGAGTCCCTG 300  
 QY 301 GCTCCCTTCTCGGATAGTGTCAAGAGTCAAAATACGTGCAAGGACCAACCCATCT 360  
 Db 301 GCTCCCTTCTCGGATAGTGTCAAGAGTCAAAATACGTGCAAGGACCAACCCATCT 360  
 QY 361 GGCGGGGGCAATGTCTCTATTACCCAGAGTCCCTCCCTACTACCGCTGTGTGTAAACAC 420  
 Db 361 GGCGGGGGCAATGTCTCTATTACCCAGAGTCCCTCCCTACTACCGCTGTGTGTAAACAC 420

QY 421 CTTTACACAGGTCCTCCAGTGTCTCCCAAGTGGTTCTCTATGACGAGCAAAACCCCTGCCAG 480  
 Db 421 CTTTACACAGGTCCTCCAGTGTCTCCCAAGTGGTTCTCTATGACGAGCAAAACCCCTGCCAG 480  
 QY 481 AATGGGGTACCTGCTCCGGGATAGCGGATCCAGTTCACCTGTGCTGCTCCCGAC 540  
 Db 481 AATGGGGTACCTGCTCCGGGATAGCGGATCCAGTTCACCTGTGCTGCTCCCGAC 540  
 QY 541 CAGTTCAAGGGGAAATTCGTGTAATAGTTCCTGATGCTGCTATGTTGGCGATGCTAC 600  
 Db 541 CAGTTCAAGGGGAAATTCGTGTAATAGTTCCTGATGCTGCTATGTTGGCGATGCTAC 600  
 QY 601 TCTTACCGAGGAAATAGTAGGACAGTCAACGACATGCTGCTTACTTGGAACTCC 660  
 Db 601 TCTTACCGAGGAAATAGTAGGACAGTCAACGACATGCTGCTTACTTGGAACTCC 660  
 QY 661 CACCTCTCTTGACGAGAAATACAAATGTTTATGAGGATGCTGTAACCCATGGATT 720  
 Db 661 CACCTCTCTTGACGAGAAATACAAATGTTTATGAGGATGCTGTAACCCATGGATT 720  
 QY 721 GGGGAAACAATTTCTGCGAAACCCAGATGCGGAGCAAAAGCCCTGCTTATATAA 780  
 Db 721 GGGGAAACAATTTCTGCGAAACCCAGATGCGGAGCAAAAGCCCTGCTTATATAA 780  
 QY 781 GTTACCAATGACAAAGTGAATGCGAATACTGTGATGTTCTCAGCCTGCTCAGCCAGGAC 840  
 Db 781 GTTACCAATGACAAAGTGAATGCGAATACTGTGATGTTCTCAGCCTGCTCAGCCAGGAC 840  
 QY 841 GTTGCCTTACCCAGAGGAAAGCCCTGAGGCTCAACAGCTTCGGGGTTGACTCC 900  
 Db 841 GTTGCCTTACCCAGAGGAAAGCCCTGAGGCTCAACAGCTTCGGGGTTGACTCC 900  
 QY 901 TGTGGAAGACTGAGATAGCAGAGAGAGATCAAGAGATCTATGGAGGCTTTAAGAGC 960  
 Db 901 TGTGGAAGACTGAGATAGCAGAGAGAGATCAAGAGATCTATGGAGGCTTTAAGAGC 960  
 QY 961 ACGGCGGCAAGCACCCATGCGAGGCTCCCTCCAGCTCTGCTGCTGCTGACCATCTCC 1020  
 Db 961 ACGGCGGCAAGCACCCATGCGAGGCTCCCTCCAGCTCTGCTGCTGCTGACCATCTCC 1020  
 QY 1021 ATGCCCCAGGGCCACTCTCTGTTGGGGCGCTCATCCACCCCTGCTGGGTGCTCACTGCT 1080  
 Db 1021 ATGCCCCAGGGCCACTCTCTGTTGGGGCGCTCATCCACCCCTGCTGGGTGCTCACTGCT 1080  
 QY 1081 GCCCAGCTGACCGACATAAAGAACAGACATCTAAAGTGTGTAGGGGACAGGAGCTG 1140  
 Db 1081 GCCCAGCTGACCGACATAAAGAACAGACATCTAAAGTGTGTAGGGGACAGGAGCTG 1140  
 QY 1141 AAGAAAGAGAAATTTATGAGCAGAGCTTTAGGGTGGAGAGATATTCAAGTACAGCAC 1200  
 Db 1141 AAGAAAGAGAAATTTATGAGCAGAGCTTTAGGGTGGAGAGATATTCAAGTACAGCAC 1200  
 QY 1201 TACAATGAAGAGATGAGATTCCCAATGATATTGCAATTGCTCAAGTTAAAGCCAGTG 1260  
 Db 1201 TACAATGAAGAGATGAGATTCCCAATGATATTGCAATTGCTCAAGTTAAAGCCAGTG 1260  
 QY 1261 GATGGTCACTGTGCTTAGAATCCAAATAGTGAAGATGCTGCTGCTGCTGCTGCTGCTG 1320  
 Db 1261 GATGGTCACTGTGCTTAGAATCCAAATAGTGAAGATGCTGCTGCTGCTGCTGCTGCTG 1320  
 QY 1321 TTTCCCTCTGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
 Db 1321 TTTCCCTCTGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380  
 QY 1381 TCCCGCCAGCTCTCTGGATGCGAAAGTCAAGTGAATGCGCAACACTTTGTGCAACTCCGC 1440  
 Db 1381 TCCCGCCAGCTCTCTGGATGCGAAAGTCAAGTGAATGCGCAACACTTTGTGCAACTCCGC 1440  
 QY 1441 CAACCTATGACCACTGATGATGAGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
 Db 1441 CAACCTATGACCACTGATGATGAGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
 QY 1501 GGGCAAGACACCTGCGAGGGTGAATCTGAGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1560

Db	1501	GGGCAAGACACCTGCGAGGGTGA	1560		241	CCCTGTGAACACAGCGTGGGAC	300	
Qy	1561	TACTACGTCTATGGGATGAGCTGGGCGCTGGAGTGTGGGAAGAGCGCCAGGGTCTAC	1620		301	GCTCCTTTCTCTGCGAATAAGTGTGAGAAAGTCAAAATACGTGCAAGAGCAACCCATGT	360	
Db	1561	TACTACGTCTATGGGATGAGCTGGGCGCTGGAGTGTGGGAAGAGCGCCAGGGTCTAC	1620		301	GCTCCTTTCTCTGCGAATAAGTGTGAGAAAGTCAAAATACGTGCAAGAGCAACCCATGT	360	
Qy	1621	ACCCAAAGTTACCAAAATTCCTGAATTCGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC	1680		361	GGCGGGGCGCAATGTCTCTATTACCCAGAGTCTCCCTACTACCGCTGTGTCTGTAACAC	420	
Db	1621	ACCCAAAGTTACCAAAATTCCTGAATTCGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC	1680		361	GGCGGGGCGCAATGTCTCTATTACCCAGAGTCTCCCTACTACCGCTGTGTCTGTAACAC	420	
Qy	1681	TAA 1683			421	CCTTACACAGTCCAGCTCTCCCAAGTGGTTCCTGTATGACAGGCAAAACCCCTGCCAG	480	
Db	1681	TAA 1683			421	CCTTACACAGTCCAGCTCTCCCAAGTGGTTCCTGTATGACAGGCAAAACCCCTGCCAG	480	
RESULT 7								
US-10-391-215-4								
; Sequence 4, Application US/10391215								
; Publication No. US2004009543A1								
; GENERAL INFORMATION:								
; APPLICANT: KIEHL, STEFAN								
; APPLICANT: WILLEIT, JOHANN								
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF								
; APPLICANT: REIMISCH, JUERGEN								
; APPLICANT: WEIMER, THOMAS								
; APPLICANT: FEUSSNER, ANNETTE								
; APPLICANT: STOEHR, HANS-ARNOLD								
; APPLICANT: DOERSAM, VOLKER								
; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE								
; TITLE OF INVENTION: (FSAP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS								
; FILE REFERENCE: 06478.1457-01								
; CURRENT APPLICATION NUMBER: US/10/391,215								
; CURRENT FILING DATE: 2003-03-19								
; PRIOR APPLICATION NUMBER: 09/912,559								
; PRIOR FILING DATE: 2001-07-26								
; PRIOR APPLICATION NUMBER: DE 100 36 641.4								
; PRIOR FILING DATE: 2000-07-26								
; PRIOR APPLICATION NUMBER: DE 100 50 040.4								
; PRIOR FILING DATE: 2000-10-10								
; PRIOR APPLICATION NUMBER: DE 100 52 319.6								
; PRIOR FILING DATE: 2000-10-21								
; PRIOR APPLICATION NUMBER: DE 101 18 706.8								
; PRIOR FILING DATE: 2001-04-12								
; NUMBER OF SEQ ID NOS: 8								
; SOFTWARE: Patentin Ver. 2.1								
; SEQ ID NO 4								
; LENGTH: 1683								
; TYPE: DNA								
; ORGANISM: Homo sapiens								
US-10-391-215-4								
Query Match								
Best Local Similarity 99.8%; Score 1679.8; DB 16; Length 1683;								
Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;								
Qy	1	ATGTTTGCAGGATGCTCGATCCATGTTCTCGTGTATGCTGTTAATGGTCTGTGGGAAAGACA	60		1081	GCCCACTGCACCGACATAAAACACAGACATCTAAAGGTGGTGTAGGGGACCCAGGACCTG	1140	
Db	1	ATGTTTGCAGGATGCTCGATCCATGTTCTCGTGTATGCTGTTAATGGTCTGTGGGAAAGACA	60		1081	GCCCACTGCACCGACATAAAACACAGACATCTAAAGGTGGTGTAGGGGACCCAGGACCTG	1140	
Qy	61	GCCTGTGGTCTCCCTGATGCTCTTATTGGAAGCCCTGGACCCAGACTGGACCCCTGAC	120		1141	AAGAAGAAGAAATTTTCATGACAGAGCTTTAGGGTGGGAAGATATTTCAAGTACAGCCAC	1200	
Db	61	GCCTGTGGTCTCCCTGATGCTCTTATTGGAAGCCCTGGACCCAGACTGGACCCCTGAC	120		1141	AAGAAGAAGAAATTTTCATGACAGAGCTTTAGGGTGGGAAGATATTTCAAGTACAGCCAC	1200	
Qy	121	CAGTATGATTACAGCTACGAGGATTATAATCAGGAAGAGACACACAGTAGCACACTTACC	180		1201	TACATGAAGAGATGAGATTCGCCACATGATATTGCAATGCTCAAGTTAAAGCCAGTG	1260	
Db	121	CAGTATGATTACAGCTACGAGGATTATAATCAGGAAGAGACACACAGTAGCACACTTACC	180		1201	TACATGAAGAGATGAGATTCGCCACATGATATTGCAATGCTCAAGTTAAAGCCAGTG	1260	
Qy	181	CATGCTGAGATCCCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC	240		1261	GATGGTCACTGTGCTTAGAATCCAAATACGTTGAAGAGCTGTGTGCTTGCCTGATGGGTCC	1320	
Db	181	CATGCTGAGATCCCTGACTGCTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC	240		1261	GATGGTCACTGTGCTTAGAATCCAAATACGTTGAAGAGCTGTGTGCTTGCCTGATGGGTCC	1320	
Qy	241	CCCTGTGAACACCGTGGGAGTGCCTGCTCCATGGGAGCACCTTCACATGACGTGCGTG	300		1321	TTTCCCTCTGGGAGTGAAGTGCACATCTCTGCTGGGTGTTTACAGAAACAGGAAAGGG	1380	



```
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Rannmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11164
; LENGTH: 451
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006097.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
US-09-864-761-11164

Query Match      7.1%; Score 120.2; DB 9; Length 451;
Best Local Similarity 97.6%; Pred. No. 8.6e-29;
Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      324 TCAGAAAGTCAAAATACGTGCAAGGACACCCCATGTGGCGGGGCAATGTCTCATTTAC 383
Db      451 TCCTACAGTCAAAATACGTGCAAGGACACCCCATGTGGCGGGGCAATGTCTCATTTAC 392
Qy      384 CCAGAGTCTCTCTACTACCGCTGTGTGTAAACACCCCTTACACAGGTCCACAGTGTCTC 443
Db      391 CCAGAGTCTCTCTACTACCGCTGTGTGTAAACACCCCTTACACAGGTCCACAGTGTCTC 332
Qy      444 CCAAG 448
Db      331 CCAAG 327

RESULT 11
US-09-954-456-552
; Sequence 552, Application US/09954456
; Patent No. US20020115057A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 552
; LENGTH: 2036
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-552

Query Match      7.0%; Score 117.6; DB 9; Length 2036;
Best Local Similarity 49.8%; Pred. No. 1.5e-27;
Matches 507; Conservative 0; Mismatches 559; Indels 54; Gaps 10;

Qy      469 AACCCCTGCCAGATGGGGCTACCTGCTCCGGGCATAGGGAGATCCAGTTCCACCTGT 528
Db      745 AGCCCTTGCTGAACGGGGGCACCTGCCACCTGTCGGCCACCGGACCCACCTGTGT 804
Qy      529 GCCTGTCGGACCAAGTTTCAAGGGGAATTCGTCAAAATAGGTTCTGATGA---CTGCTAT 585
Db      805 GCCTGCCACCAAGGCTTGGTGGAGGCTCTGCAACATCAGAGCTGTATGACGCTGTTC 864
Qy      586 GTTGGCATGCTACTCTTACCGAGGGAAAATGAATAGGACAGTCAACACGATGCGTGC 645
Db      865 TTGGGAAACGGCACTGGGTACCGTGGCGTGGCCAGCACCTCAGCCTCGGGCTCAGCTGC 924
Qy      646 CTTTACTGGAATCCCACTCTCTTTCAGGAGAAATTACAACATGTTTATGAGGATGCT 705
Db      925 CTGGCTGGAATCTCGATCTCTTACAGGAGTGCACGTGGACTCGTGGGCGCGCG 984
Qy      706 GAAACCCATGGGATTTGGGGAACACAATTTCTGCAGAAACCCAGATGCGGACAAAAGCCC 765
Db      985 GCGCTGCTGGGCTGGGCGCCCATGCTACTTGCAGGAATCGGACAATGACGAGAGGCC 1044
Qy      766 TGGTCTTTTAAAGTTACCAATGCAAGGTGAATGGGNAATCTGTGATGTCTCAGCC 825
Db      1045 TGGTCTTACGT---GGTGAAGACAGCGCTCTCTCTGGGAGTACTCTCCGCTGGAGGCC 1101
Qy      826 TGCTCAGCCCGAGGAGTGTGCTTACCCAGAGGAAAGCCCCACTGAGCCATCAACCAAGCTT 885
Db      1102 TGGGAATCCCT-----CACCAGAGTCCAACCTGTCAACCGGATCTCTCTGGCACCCTG 1152
Qy      886 CCGGGGTTTGACTCTCTGTGGAAGNCTGAGTAGCAGAGAGAGATCAAGAGATCTAT 945
Db      1153 CTTGAGCCAGCCTCTCCCGGGCGCGAGGCTGTGCGGAGGAGGCAAGAGAGACGTTTC 1212
Qy      946 GGAGGCTTTAAGAGCACCGCGGGAAGCACCCATGCGCAGGGGCTCCCTCCAGTCTCTCGCTG 1005
```

Db 1213 CTGGCGCACGATATCATCGCGGGCTCTCTCTGCTGCGCGGGTGCACCCCTG----- 1265  
Qy 1006 CTTCTGACCATCTCATCCCGAGGCGACCTTCTGTGTGGGGCTGATCCACCCCTGCG 1065  
Db 1266 --GCTGGCGGCCATCTACATCGGGACAGCTTCTGCGCGGAGCGCTGCTCCACACCTGC 1323  
Qy 1066 TGGGTGCTCACTGCTGCCACATGCGACGACATATAAAGACAGACATCTAA-----AGGTG 1119  
Db 1324 TGGGTGGTGTGGCGCGCCACCTGCTTCTCCACAGCCCGCCAGGACAGCGTCTCGGTG 1383  
Qy 1120 GTGTAGGGGACAGGACCTGAAGAAAGAAATTTTCATGACGAGCTTTAGGGTGGAG 1179  
Db 1384 GTGTGGGCGACGACTTCTTCAACGCGACGCGGACGCTGACGACGCTTGGCATCGAG 1443  
Qy 1180 AGATATTCAAGTACAGCCTACATGAAGAGATGAGATTCGCCCAATGATATTGCA 1239  
Db 1444 AAGTACATCCCGTACACCCCTGCTCGGGTTTCAACCCCGAGCGAC--GACCTGTC 1500  
Qy 1240 TTGCTCAAGTTAAAGCCAGTGATGGTCACTGTGCTCTAGAAATCCAAATACGTGAAGACT 1299  
Db 1501 CTGATCCGGCTGAAGAAAGGAGCGCTGTGCGCACGCTCGCAGTTCTGTGAGGCC 1560  
Qy 1300 GTGTGCTTGCCTGAT-----GGGTCTTTTCCCTCTGGGAGTGAAGTGCACATCTTGGC 1353  
Db 1561 ATCTGCTTGCCTGAGCGCGGAGCGCTTCCCGGAGGACCAAGTGCAGATTCGCGGC 1620  
Qy 1354 TGGGG-----TGTTACAGAAACAGAAAGGGTCCCGCAGCTCTCTGATGCGCAAGTC 1407  
Db 1621 TGGGCCACTTGGATGAGAACTGAGCGGCTACTCCAGTCCCTGCGGAGGCGCTTGGTC 1680  
Qy 1408 AAGCTGATTGCAACACTTTGCAACTCCCGCCAACTCTATGACCAATGATGATGAC 1467  
Db 1681 CCCCTGGTGGCGGACCAAGTGCAGAGCCCTGAGGCTTACCGGCGCGACATCAGCCCC 1740  
Qy 1468 AGTATGATCTGTGAGGAACTTTCAGAACTTGGGCAAGACCTGCGGAGGCTGACCTCT 1527  
Db 1741 AACTGCTCTGTGCGGCTACTT---CGACTGAAAGTCCGAGCGCTGCGGAGGACTCA 1797  
Qy 1528 GGAGGCCCCCTGACCTGTGAGAGGACGCGCACTACTAGTCTATGGGATAGTGAAGTGG 1587  
Db 1798 GGGGGCCCCCTGGCTGCGAGAAAGCGGCTGAGGCTTACCTTACGCGCATCATCAGCTGG 1857  
Qy 1588 GCCTGGAGTGGG-----AGAGGCGGAGGGTCTACCCCAAGTTACCAATTCCTG 1641  
Db 1858 GGTGACGGCTGCGGGCGGCTCCACAGCGGGGGTCTACCCCGCTGGCCAACTATGTG 1917  
Qy 1642 AATTGGATCAAGCCACCAT 1661  
Db 1918 GACTGGATCAACGACCGGAT 1937

RESULT 12  
US-09-880-107-1612  
; Sequence 1612, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; PRIOR FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1612  
; LENGTH: 2036  
; TYPE: DNA

; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D14012  
US-09-880-107-1612  
  
Query Match 7.0%; Score 117.6; DB 9; Length 2036;  
Best Local Similarity 49.8%; Pred. No. 1.5e-27;  
Matches 607; Conservative 0; Mismatches 559; Indels 54; Gaps 10;  
  
Qy 469 AACCCCTGCCAGATGGGGTACCTCTCCCGCATTAAGCGGAGATCCAAAGTTCACTGT 528  
Db 745 AGCCCTTGCTGAACGGGGGACCTGCCACCTGATCGTGGCCACCGGGACCACTGTGTG 804  
Qy 529 GCCTGTCCCGACAGTTCAAGGGGAATTTGTGAATAGGTTCTGATGA---CTGCTAT 585  
Db 805 GCCTGCCACAGGCTTGGTGAACGGCTCTGCAACTCGAGCTGATGAGCGTCTTC 864  
Qy 586 GTTGGCGATGGCTACTTTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCGTGC 645  
Db 865 TTGGGAAACGGCACTGGGTACCGTGGCTGGCCAGCACCTCAGCTCGGGCCTCAGCTGC 924  
Qy 646 CTTTACTGNACTCCACCTCTCTTGCAGGGAATTAACAATGTTTATGAGGATGCT 705  
Db 925 CTGGCTGGAACCTCCGATCTGTCTTACAGGAGCTGACGTGACTTCCGTGGCGCGCG 984  
Qy 706 GAAACCATGGGATTGGGGAACACAATTTCTGCAGAAACCCAGATCGGACGAAAGGCC 765  
Db 985 GCCTCTCTGGCTTGGCGCCCATGCTACTTCCGGAATCCGACAATGACGAGAGGCC 1044  
Qy 766 TGTGCTTTATTAAGATTACCAATGAAAGTGAATGGAATGATGATGATGATGATGAT 825  
Db 1045 TGTGCTGATGCT---GGTGAAGGACAGCGGCTCTCTCTGGGAGTACTGCGCGCTGGAGGCC 1101  
Qy 826 TGCTGAGCCAGGACCTTGCCTACCCAGAGGAAAGCCCACTGAGGCATCAACCAAGCTT 885  
Db 1102 TGGATTCCT-----CACCAGTCCAACTGTCACTGCTCCTGGGACCTTG 1152  
Qy 886 CCGGGTTGACTCTGTGGAAGAAGCTGAGATGACGAGAGGAAGATCAAGAAATCTAT 945  
Db 1153 CTTGAGCAGCTTCCCGGGCGCCAGGCTTGGCGAGGAGGACCAAGAAAGGAGGCTTC 1212  
Qy 946 GGAGGCTTTAAGAGACGCGGGGCAAGCACTGAGCGGCTCTCTCTGCTGCTGCGCTCCGCTG 1005  
Db 1213 CTGGGCGACGATCATTCGGGCGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1265  
Qy 1006 CTTCTGACCATCTCCATGCCCCAGGCCACTTCTGTGTGGGGCGGTGATCCACCCCTGC 1065  
Db 1266 --GCTGGCGCCATCTACATCGGGGACAGCTTCTGCGCGGGAGCGCTGTGTCACACCTGC 1323  
Qy 1066 TGGGTGCTCACTGCTGCCACCTGAAAGAAAGAAATTTTCATGACGAGCTTTAGGGTGGAG 1179  
Db 1324 TGGGTGCTGCGCGCCCACTCTCTTCCACAGCCCCCGGAGGACGCTTCCGCTG 1383  
Qy 1120 GTCTAGGGGACCGAGGACCTGAAAGAAAGAAATTTTCATGACGAGCTTTAGGGTGGAG 1179  
Db 1384 GTCTGGGCCAGCACTTCTTCAACCGCACGAGCGTGCAGACCTTCGSCATCGAG 1443  
Qy 1180 AGATATTCAAGTACAGCCACTACATGAAGAGATGAGATTCGCCCAATGATATTGCA 1239  
Db 1444 AAGTACATCCCGTACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500  
Qy 1240 TTGCTCAAGTTAAAGCCAGTGATGGTCACTGTGCTTGAATTCAGAAATCCAAATACGTGAAGACT 1299  
Db 1501 CTGATCCGGCTGAAGAAAGGAGCGCTGTGCGCACGCTGCGAGTTCTGTGAGGCC 1560  
Qy 1300 GTGTGCTTGCCTGAT-----GGGTCTTTTCCCTCTGGGAGTGAAGTGCACATCTTGGC 1353  
Db 1561 ATCTGCTTGCCTGAGCGCGGAGCGCTTCCCGGAGGACCAAGTGCAGATTCGCGGC 1620  
Qy 1354 TGGGG-----TGTTACAGAAACAGAAAGGGTCCCGCAGCTCTCTGATGCGCAAGTC 1407  
Db 1621 TGGGCCACTTGGATGAGAACTGAGCGGCTACTCCAGTCCCTGCGGAGGCGCTTGGTC 1680

```
QY 1408 AAGCTGATTGCCAACACTTTGTGCAACTCCGCCCACTCTATGACCAATGATTGATGAC 1467
Db 1681 CCCCTGGTCCCGACACCAAGTGCAGCAGCCTTACGGTCTACGGCGCGCAATCAGCCCC 1740
QY 1468 AGTATGATCTGTGCAGGAATTTTCAGAAACCTTGGGCAAGACACTTGCAGGGTGACTCT 1527
Db 1741 AACATGCTCTGTGCCGGTACTT---CGACTGCAAGTCCGACGCTTGCAGGGGACTCA 1797
QY 1528 GGAAGGCCCTGACCTGTGAGAGGACGCGCACTACTAGTCTATGAGGATAGTGAGCTGG 1587
Db 1798 GGGGGGGCCCTCGCCCTGCGAGAAGACGGGTGGCTTACTTACGGCATCATCAGCTGG 1857
QY 1588 GGCTCGAGTGTGGG-----AAGAGCGCCAGGGGTCTACACCAAGTTACCAAAATTCCTG 1641
Db 1858 GGTGACGGCTGCGGGCGGCTCCACAAGCGGGGTCTACACCGCGGTGCGCCCACTATGTG 1917
QY 1642 AATTGGATCAAGCCACCAT 1661
Db 1918 GACTGGATCAACGACCGGAT 1937

RESULT 13
US-09-864-761-27791/c
; Sequence 27791, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
```

```
; SEQ ID NO 27791
; LENGTH: 117
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006037.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
; OTHER INFORMATION: NT HIT: S8182.1 EVALUE 8.00e-60
; OTHER INFORMATION: SWISSPROT HIT: Q25464, EVALUE 1.00e-03
; OTHER INFORMATION: EST HUMAN HIT: T68666.1, EVALUE 7.00e-55
US-09-864-761-27791

Query Match 7.0%; Score 117; DB 9; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.5e-28;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 TGCAAAATACGTGCAAGGACAAACCCATGTGGCGGGCCCAATGTCTCATTTACCCAGATC 391
Db 117 TGCAAAATACGTGCAAGGACAAACCCATGTGGCGGGCCCAATGTCTCATTTACCCAGATC 58
QY 392 CTCCTACTACCGCTGTGTGTAACACCCCTTACACAGCTCCAGCTGCTCCCAAG 448
Db 57 CTCCTACTACCGCTGTGTGTAACACCCCTTACACAGCTCCAGCTGCTCCCAAG 1

RESULT 14
US-10-087-192-590
; Sequence 590, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 590
; LENGTH: 1302
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-087-192-590

Query Match 5.5%; Score 93.2; DB 13; Length 1302;
Best Local Similarity 49.7%; Pred. No. 1.6e-19;
Matches 586; Conservative 0; Mismatches 523; Indels 69; Gaps 11;

QY 526 TGTGCTGTCCCGACCACTTCAAGGGAAATTTCTGTAATAGGT---TCTGATGACTGC 582
Db 154 TGCAGCTGCCCAAGGAAATTCAGGGGAGCACTGTGAGATAGATGCATCAAAAACCTGC 213
QY 583 TATGTTGGGATGGCTACTCTTACGAGGAAATGATAGGACAGTCAACACGATCGC 642
Db 214 TATCATGGAATGGTGTACTTTACGAGGAAAGGCCAACACTGATACCAAGGTGGCCC 273
QY 643 TGCCTTTTACTGGAATCCCACTCTCTTTGAGGAAATTAACAATTTTATGAGGAT 702
Db 274 TGCCTGGCTTGAATGCGCTGCTGCTCTTCCAGAAACCTACATGCCCCACAGACTGAT 333
QY 703 GCTGAAACCCATGGGATTTGGGGAACACATTTCTGCAGAAACCCAGATGCGGCGAAAG 762
Db 334 GCTATTAGCTTAGGCTGGGGAACACATTTACTGAGGAACCTTGACACACCAAGGCA 393
```





```
QY 1297 ACTGTGTGCTGCTGATGGTCCCTT-----TCCCTCTGGGAGTGAGTGCCACATCTCT 1350
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 953 ACCATCTGCTGCCCCCAAGGTTTACTGATGCTCCGTTTGGTTTCAGACTGTGAGATCACT 1012
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1351 GGCTGGGGT---GTTACAGAAACAGGAAAGGGTCCCGCAGCTCCTGGATG---CCAAA 1404
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1013 GGGTTTGGAAAAGAGTCTGAAAGTGACTTCTCTATCCAAAGNACCTGAAATGTCGCTC 1072
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1405 GTCAGCTGATGCCAACACTTTGTGCAACTCCCGCAACTCTATGACCACATGATTGAT 1464
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1073 GTAAAGCTTGTCTCATGAACAGTGTATGCAGCCCCCACTACTATGGCTCTGAAATTAAT 1132
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1465 GACAGTATGATCTGTGCAGGAAATCTTCAGAAACCTGGGCAAGACACCTGCCAGGGTGAC 1524
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1133 TATAAATGCTGTGCTGCGACCCAGAGTGAATAAC---AGATTCTTGCAGGGCGAT 1189
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1525 TCTGGAGGCCCTGACCTGTGAGAGGACGGCACCTACTAGCTCTATGGATAGTAGC 1584
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1190 TCTGGAGGCCGCTTATCTGTAACATCGAAGGCCGCCCAACTCTGAGTGGGATTGTGAGC 1249
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1585 TGGGGCTGGAGTGTG-----GGAAGAGCCAGGGGTCTACACCCCAAGTTACCAATTC 1638
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1250 TGGGGCCGAGGATGTGAGAGAAACACAGCCCGGTGTCTACACAGGGTCTCACACTTC 1309
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1639 CTGAATTGGATCAAAGCCACCATCAAAAAGTGAAGTGG 1676
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1310 CTGACTGGATTCAATCCCAATTGGAGAGAGAAAGG 1347
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: May 25, 2004, 13:22:31  
Job time : 538 secs



```

detection
JOURNAL: parent: EP 1182258-A 2 27-FEB-2002;
FEATURES: Beihing GmbH (DE)
source: 1..1693
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"

```

## ORIGIN

Query Match	100.0%;	Score 1683;	DB 6;	Length 1683;
Best Local Similarity	100.0%;	Prod. No. 0;		
Matches 1683;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGTTTGCCAGGATGCTGATCTCCATGTTCTGCTGTTTAAATGCGTCTGCGGAAAGACA	60	
DB	1	ATGTTTGCCAGGATGCTGATCTCCATGTTCTGCTGTTTAAATGCGTCTGCGGAAAGACA	60	
QY	61	GCTGTGGGTTCTCCCTGATGTCCTTTATTTGAAAGCCTGGACCCAGATCGGACCCCTGAC	120	
DB	61	GCTGTGGGTTCTCCCTGATGTCCTTTATTTGAAAGCCTGGACCCAGATCGGACCCCTGAC	120	
QY	121	CAGTATGATTACAGCTACGAGGATTATTAATCAGAAGAGAGAACACCAAGTAGCACACTTACC	180	
DB	121	CAGTATGATTACAGCTACGAGGATTATTAATCAGGAGAGAGAACCAAGTAGCACACTTACC	180	
QY	181	CATGTGGAATCTGACTGCTGCTACTCAGGACCAAGCTGATCCATGCCAGCCCAAC	240	
DB	181	CATGTGGAATCTGACTGCTGCTACTCAGGACCAAGCTGATCCATGCCAGCCCAAC	240	
QY	241	CCCTGTGAACACGGTGGGAGCTGCTCGTCCATCGGAGCACCTTCACATCGACGCTGCCGTG	300	
DB	241	CCCTGTGAACACGGTGGGAGCTGCTCGTCCATCGGAGCACCTTCACATCGACGCTGCCGTG	300	
QY	301	GCTCCTTTCTTGGGAATAAGTGTGAGAAAGTCAGAAATACGTGCAAGGACCAACCCATGT	360	
DB	301	GCTCCTTTCTTGGGAATAAGTGTGAGAAAGTCAGAAATACGTGCAAGGACCAACCCATGT	360	
QY	361	GCGCGGGCCATGTCCTATTACCCAGAGTCCTCCCTACTACCGCTGTGTCTGTAACAC	420	
DB	361	GCGCGGGCCATGTCCTATTACCCAGAGTCCTCCCTACTACCGCTGTGTCTGTAACAC	420	
QY	421	CCTTACACAGTCCCAGCTGCTCCCAAGTGGTTCTCTGTATGCAAGGCCAAACCCCTGCCAG	480	
DB	421	CCTTACACAGTCCCAGCTGCTCCCAAGTGGTTCTCTGTATGCAAGGCCAAACCCCTGCCAG	480	
QY	481	AATGGGCTACTGCTCCGGCATAGCGGAGATCCAAAGTTCACTGTGCTGTCCGGAC	540	
DB	481	AATGGGCTACTGCTCCGGCATAGCGGAGATCCAAAGTTCACTGTGCTGTCCGGAC	540	
QY	541	CAGTTCAAGGGGAAATTCGTGAAATAGGTTCTCATGACTGCTATGTTGGCGATGGCTAC	600	
DB	541	CAGTTCAAGGGGAAATTCGTGAAATAGGTTCTCATGACTGCTATGTTGGCGATGGCTAC	600	
QY	601	TCTTACCGAGGAAATGAATAGNACAGTCAACAGCATCGGTGCTTTACTTGAATCC	660	
DB	601	TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATCGGTGCTTTACTTGAATCC	660	
QY	661	CACCTCCTCTTCAGGAGAAATTACAACATGTTTATGGAGGATGCTGAAACCCATGGGATT	720	
DB	661	CACCTCCTCTTCAGGAGAAATTACAACATGTTTATGGAGGATGCTGAAACCCATGGGATT	720	
QY	721	GGGGAACAAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGGTCTTTTATAA	780	
DB	721	GGGGAACAAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGGTCTTTTATAA	780	
QY	781	GTTTACCAATGCAAGGTGAATGGGAATACGTGATGTCTCAGCTGCTCAGGCCAGGAC	840	
DB	781	GTTTACCAATGCAAGGTGAATGGGAATACGTGATGTCTCAGCTGCTCAGGCCAGGAC	840	
QY	841	GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACAAAGCTTCGGGGTTTGTACTCC	900	
DB	841	GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACAAAGCTTCGGGGTTTGTACTCC	900	

Qy	901	TGTGGAAGACTGAGATAGCAGAGAGAAAGATCAATGAGAGATCTATGAGAGCTTTAAGAGC	960
Db	901	TGTGGAAGAAGCTGAGATAGCAGAGAGAAAGATCAAGAGAAATCTATGAGAGCTTTAAGAGC	960
Qy	961	ACGGCGGGCAAGACACCCATGGCAGGGCGTCCCTCCAGTCCTCGTGCTCCCTCTGAGCATCTCC	1020
Db	961	ACGGCGGGCAAGACACCCATGGCAGGGCGTCCCTCCAGTCCTCGTGCTCTGAGCCATCTCC	1020
Qy	1021	ATGCCCCAGGGCCACTTCTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCATCTGCT	1080
Db	1021	ATGCCCCAGGGCCACTTCTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCATCTGCT	1080
Qy	1081	GCCCACTGCACCGACATATAAACAACAGACATCTAAAGTGTGCTAGGGGACCAAGGACCTG	1140
Db	1081	GCCCACTGCACCGACATATAAACAACAGACATCTAAAGTGTGCTAGGGGACCAAGGACCTG	1140
Qy	1141	AAGAAAGAAGAAATTTCTATGAGCAGAGCTTTAGGTGTGAGAGATATTCAAGTACAGCCAC	1200
Db	1141	AAGAAAGAAGAAATTTCTATGAGCAGAGCTTTAGGTGTGAGAGATATTCAAGTACAGCCAC	1200
Qy	1201	TACAATGAAGAGATGAGATTTCCCACAATGATATTGCCATGTCTCAAGTTAAAGCCAGTG	1260
Db	1201	TACAATGAAGAGATGAGATTTCCCACAATGATATTGCCATGTCTCAAGTTAAAGCCAGTG	1260
Qy	1261	GATGGTCACGTGTGCTCTAGAATCCAAATACGTGAAGACTGTGTGCTTGCTGATGGGTCC	1320
Db	1261	GATGGTCACGTGTGCTCTAGAATCCAAATACGTGAAGACTGTGTGCTTGCTGATGGGTCC	1320
Qy	1321	TTTCCCTCTGGAGTGAGTGCCACATCTCTGGCTGGGGTGTACAGAAAACAGAAAAGGG	1380
Db	1321	TTTCCCTCTGGAGTGAGTGCCACATCTCTGGCTGGGGTGTACAGAAAACAGAAAAGGG	1380
Qy	1381	TCCCGCAGCTCTGGATGTCCAAGTCAAGCTGATTCGCCAACACTTTGTGCACTCCCGC	1440
Db	1381	TCCCGCAGCTCTGGATGTCCAAGTCAAGCTGATTCGCCAACACTTTGTGCACTCCCGC	1440
Qy	1441	CAACTCTATGACCAACATGATTGATGACAGTATGATCTGTGCAGAAATCTTCAGAAACCT	1500
Db	1441	CAACTCTATGACCAACATGATTGATGACAGTATGATCTGTGCAGAAATCTTCAGAAACCT	1500
Qy	1501	GGGCAAGACACTGCCAGGTGACTCTGGAGGCCCTTGACCTGTGAGAAAGGACGGCACC	1560
Db	1501	GGGCAAGACACTGCCAGGTGACTCTGGAGGCCCTTGACCTGTGAGAAAGGACGGCACC	1560
Qy	1561	TACTACGTCTATGGGATAGTGAGCTGGGGCTGGAGTGTGAGAAAGGCCAGGGGTCTAC	1620
Db	1561	TACTACGTCTATGGGATAGTGAGCTGGGGCTGGAGTGTGAGAAAGGCCAGGGGTCTAC	1620
Qy	1621	ACCCAAGTTACCAAAATTCCTGAAATTTGGATCAAAGCCACCATCAAAAGTGAAGTGGCTTC	1680
Db	1621	ACCCAAGTTACCAAAATTCCTGAAATTTGGATCAAAGCCACCATCAAAAGTGAAGTGGCTTC	1680
Qy	1681	TAA 1683	
Db	1681	TAA 1683	

2. T. J. L. S. E. A.

RESUBMIT 2	BD177603	1683 bp	DNA	linear	PAT 16-APR-2003
LOCUS	BD177603				
DEFINITION	Detection method with the use of factor VII activating protease mutant and specific antibody.				
ACCESSION	BD177603				
VERSION	BD177603.1	GI:30014865			
KEYWORDS	JP 2002291486-A/2.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1683)				
AUTHORS	Roemisch, J., Stoeck, H.A., Feussner, A., Lang, W., Weimer, T., Becker, M., Nerlich, C. and Naumann, G.M.				

TITLE Detection method with the use of factor VII activating protease  
 JOURNAL mutant and specific antibody  
 Patent: JP 2002291486-A 2 08-OCT-2002;  
 COMMENT AVENTIS BEHRING GMBH  
 OS Homo sapiens (human)  
 PN JP 2002291486-A/2  
 PD 08-OCT-2002  
 PR 25-JUL-2001 JP 2001224423  
 PR 26-JUL-2000 DE 10036641.4, 10-OCT-2000 DE 10050040.4 PR  
 21-OCT-2000 DE 10052319.6, 12-APR-2001 DE 10118706.8 PI  
 JUEGEN ROEMISCH, HANS ARNOLD STOEHR, ANNETTE FEUSSNER, WIEGAND PI  
 LANG,  
 PI THOMAS WEIMER, MARGRET BECKER, CLAUDIA NERLICH, GUDRUN MUTH PI  
 NAUMANN  
 PC C12N15/09, C07K16/40, C12N9/50, C12Q1/58, G01N33/53, G01N33/533, PC  
 G01N33/534  
 PC G01N33/535//C12P21/08, C12N15/00  
 CC Detection method with the use of factor VII activating CC  
 protease mutant and  
 CC specific antibody  
 FH Key Location/Qualifiers  
 FT source 1..1683  
 FT Location/Qualifiers  
 1..1683  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 1683; DB 6; Length 1683;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTTGGCCAGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGTTGGGAAAGACA 60  
 DB 1 ATGTTTGGCCAGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGTTGGGAAAGACA 60

QY 61 GCCTGTGGGTTCTCCCTGATGCTTTATTTGAAAGCCCTGACCCAGACTGACCCCTGAC 120  
 DB 61 GCCTGTGGGTTCTCCCTGATGCTTTATTTGAAAGCCCTGACCCAGACTGACCCCTGAC 120

QY 121 CAGTATGATTACAGTACAGGATTTATTAATCAGAAAGAGAACACAGTAGCACACTTACC 180  
 DB 121 CAGTATGATTACAGTACAGGATTTATTAATCAGAAAGAGAACACACAGTAGCACACTTACC 180

QY 181 CATGCTGAGATTCCTGATGCTGATCTACACTGAGGACCAAGCTGATCCATGCCGCCAAC 240  
 DB 181 CATGCTGAGATTCCTGATGCTGATCTACACTGAGGACCAAGCTGATCCATGCCGCCAAC 240

QY 241 CCCTGTGAACACGGTGGGACTGCTGCTCCATGGGACACCTTCACATGCGAGCTGCCTG 300  
 DB 241 CCCTGTGAACACGGTGGGACTGCTGCTCCATGGGACACCTTCACATGCGAGCTGCCTG 300

QY 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAAAATACGTGCAAGGACCAACCCATGT 360  
 DB 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAAAATACGTGCAAGGACCAACCCATGT 360

QY 361 GSCCGGGCCATGCTCTATACCCAGAGTCTCCCTACTACCGCTGCTGTGTAAACAC 420  
 DB 361 GSCCGGGCCATGCTCTATACCCAGAGTCTCCCTACTACCGCTGCTGTGTAAACAC 420

QY 421 COTTTACAGGTCCTCAGCTGCTCCCAAGTGTGTTCTGTATGAGGCCAAACCCCTGCGAG 480  
 DB 421 COTTTACAGGTCCTCAGCTGCTCCCAAGTGTGTTCTGTATGAGGCCAAACCCCTGCGAG 480

QY 481 AATGGGGTACTGCTCCGGGATTAAGCGGAGATCCAAAGTTCACTGTGCTGCTCCGAC 540  
 DB 481 AATGGGGTACTGCTCCGGGATTAAGCGGAGATCCAAAGTTCACTGTGCTGCTCCGAC 540

QY 541 CAGTTCAAGGGGAAATTTCTGTGAATAGTCTGATGATGCTGATGTTGGGGATGGGTAC 600  
 DB 541 CAGTTCAAGGGGAAATTTCTGTGAATAGTCTGATGATGCTGATGTTGGGGATGGGTAC 600

QY 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGGTGCCTTTACTGGAACTCC 660  
 DB 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGGTGCCTTTACTGGAACTCC 660

QY 661 CACCTCCTCTTTCAGGAGAAATTACAACTGTTTATGAGGATGCTGAAACCCATGGGATT 720  
 DB 661 CACCTCCTCTTTCAGGAGAAATTACAACTGTTTATGAGGATGCTGAAACCCATGGGATT 720

QY 721 GGGGAACACATTTCTGAGAAACCCAGATCGGACGAAAGCCCTGCTGCTTTATTTAAA 780  
 DB 721 GGGGAACACATTTCTGAGAAACCCAGATCGGACGAAAGCCCTGCTGCTTTATTTAAA 780

QY 781 GTTACCAATGACAAAGTGAATGGAAATCTGTGATGTCTCAGCCCTGCTCAGCCAGGAC 840  
 DB 781 GTTACCAATGACAAAGTGAATGGAAATCTGTGATGTCTCAGCCCTGCTCAGCCAGGAC 840

QY 841 GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTGACTCC 900  
 DB 841 GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTTGACTCC 900

QY 901 TGTGGAAGACTGAGTAGCAGAGAGAGATCAAGAGATCTATGAGGCTTTTAAGAGC 960  
 DB 901 TGTGGAAGACTGAGTAGCAGAGAGAGATCAAGAGATCTATGAGGCTTTTAAGAGC 960

QY 961 ACGGGGGCAAGCACCACATGCGAGGCTCCCTCCAGTCTCTGCTGCTGACCATCTCC 1020  
 DB 961 ACGGGGGCAAGCACCACATGCGAGGCTCCCTCCAGTCTCTGCTGCTGACCATCTCC 1020

QY 1021 ATGCCCCAGGGCCACTTCTGTGGTGGGGCTGATCCACCCCTGCTGGGTCTCACTGCT 1080  
 DB 1021 ATGCCCCAGGGCCACTTCTGTGGTGGGGCTGATCCACCCCTGCTGGGTCTCACTGCT 1080

QY 1081 GCCCACTGCACCGACATAAAAAACAGACATCTAAAGTGGTGTCTAGGGGACAGACCTG 1140  
 DB 1081 GCCCACTGCACCGACATAAAAAACAGACATCTAAAGTGGTGTCTAGGGGACAGACCTG 1140

QY 1141 AAGAAAGAGATTTTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAGTACAGCCAC 1200  
 DB 1141 AAGAAAGAGATTTTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAGTACAGCCAC 1200

QY 1201 TACAATGAAGAGATGAGATTTCCCAACATGATATTGCAATGCTCAAGTTAAAGCCAGTG 1260  
 DB 1201 TACAATGAAGAGATGAGATTTCCCAACATGATATTGCAATGCTCAAGTTAAAGCCAGTG 1260

QY 1261 GATGTCATGCTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGTCC 1320  
 DB 1261 GATGTCATGCTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGTCC 1320

QY 1321 TTTCCCTCTGGGAGTGAGTGCCACATCTCTGCTGGGGTGTTCAGAAACAGGAAAGGG 1380  
 DB 1321 TTTCCCTCTGGGAGTGAGTGCCACATCTCTGCTGGGGTGTTCAGAAACAGGAAAGGG 1380

QY 1381 TCCCGCCAGCTCTGGATGCCAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC 1440  
 DB 1381 TCCCGCCAGCTCTGGATGCCAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC 1440

QY 1441 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500  
 DB 1441 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500

QY 1501 GGGCAAGACACTGCGCAGGGTGACTGTGAGGCCCTTGAACCTGTGAGAAAGAGCGGAC 1560  
 DB 1501 GGGCAAGACACTGCGCAGGGTGACTGTGAGGCCCTTGAACCTGTGAGAAAGAGCGGAC 1560

QY 1561 TACTAGCTCTATGGGATGAGCTGGGGCTTGGAGTGAGAGAGGCGGGTCTAC 1620  
 DB 1561 TACTAGCTCTATGGGATGAGCTGGGGCTTGGAGTGAGAGAGGCGGGTCTAC 1620

QY 1621 ACCCAAGTTACCAAAATTCCTGAATTTGGATCAAGGCCACCATCAAAAGTGAAGTGGCTTC 1680  
 DB 1621 ACCCAAGTTACCAAAATTCCTGAATTTGGATCAAGGCCACCATCAAAAGTGAAGTGGCTTC 1680

QY 1681 TAA 1683  
Db 1681 TAA 1683

RESULT 3  
AX383954  
LOCUS AX383954 1683 bp DNA linear PAT 19-MAR-2002  
DEFINITION Sequence 1 from Patent EP1182258.  
ACCESSION AX383954  
VERSION AX383954.1 GI:19577525  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Roemisch, J., Stoeck, H.A., Feussner, A., Lang, W., Weimer, T.,  
Becker, M., Nerlich, C. and Much-Naumann, G.  
TITLE Mutants of the factor vll activating protease and methods for their  
detection  
JOURNAL Patent: EP 1182258-A 1 27-FEB-2002;  
Aventis Behring GmbH (DE)  
FEATURES  
Location/Qualifiers  
1..1683  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 99.8%; Score 1679.8; DB 6; Length 1683;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 168; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGTTGCCAGGATGTCATCTCCATGTTCTCTGTTAATGCTCTGTGGGAAAGACA 60  
Db 1 ATGTTGCCAGGATGTCATCTCCATGTTCTCTGTTAATGCTCTGTGGGAAAGACA 60

QY 61 GCCTGTGGGTCTCCCTGATGTTCTTTATGGAAAGCTGGACCCAGACTGGACCCCTGAC 120  
Db 61 GCCTGTGGGTCTCCCTGATGTTCTTTATGGAAAGCTGGACCCAGACTGGACCCCTGAC 120

QY 121 CAGTATGATTACAGCTACGAGGATTAATACAGGAAGAGAACACAGTAGACACATTACC 180  
Db 121 CAGTATGATTACAGCTACGAGGATTAATACAGGAAGAGAACACAGTAGACACATTACC 180

QY 181 CATGCTGAGAACTCTGACTGTTACTACATGAGGACCAAGCTGATCCATGCCAGCCCAAC 240  
Db 181 CATGCTGAGAACTCTGACTGTTACTACATGAGGACCAAGCTGATCCATGCCAGCCCAAC 240

QY 241 CCCTGTCAACACGCTGGGGACTGCTCTCCATGGGAGCACCTTCAATGACGCTGCCTG 300  
Db 241 CCCTGTCAACACGCTGGGGACTGCTCTCCATGGGAGCACCTTCAATGACGCTGCCTG 300

QY 301 GCTCCTTTCTTGGGAATAGTGTGAAAGTGAATAACGTGGAAGACACCATGTT 360  
Db 301 GCTCCTTTCTTGGGAATAGTGTGAAAGTGAATAACGTGGAAGACACCATGTT 360

QY 361 GGCCGGGGCCAAATGTCATTACCCAGAGTCTCCCTACTACCGCTGTCTGTAAACAC 420  
Db 361 GGCCGGGGCCAAATGTCATTACCCAGAGTCTCCCTACTACCGCTGTCTGTAAACAC 420

QY 421 CTTTACACAGTCCCAAGTGTCTCCCAAGTGGTTCTGTATGAGGCGCAAAACCCCTGCCAG 480  
Db 421 CTTTACACAGTCCCAAGTGTCTCCCAAGTGGTTCTGTATGAGGCGCAAAACCCCTGCCAG 480

QY 481 AATGGGCTACTCTGCTCCCGCATAGCGAGATCCAAAGTTCACTGTGCTCTCCGAC 540  
Db 481 AATGGGCTACTCTGCTCCCGCATAGCGAGATCCAAAGTTCACTGTGCTCTCCGAC 540

QY 541 CAGTTCAAGGGGAAATCTGTGAATAGGTTCTGATGACTGCTATGTTGGCGATGGCTAC 600  
Db 541 CAGTTCAAGGGGAAATCTGTGAATAGGTTCTGATGACTGCTATGTTGGCGATGGCTAC 600

QY 601 TCTTACCGAGGAAATGAATAGACAGTCAACACAGCATCGTGCCTTTACTGGAATCC 660  
Db 601 TCTTACCGAGGAAATGAATAGACAGTCAACACAGCATCGTGCCTTTACTGGAATCC 660

QY 661 CACCTCTCTTGCAGGAGAAATTAACAATGTTTATGAGGATGCTGAAAACCCATGGGATT 720  
Db 661 CACCTCTCTTGCAGGAGAAATTAACAATGTTTATGAGGATGCTGAAAACCCATGGGATT 720

QY 721 GGGGAACACAAATTTCTGCAGAAACCCAGATCGGAGGAAAGCCCTGCTTTATTTAAA 780  
Db 721 GGGGAACACAAATTTCTGCAGAAACCCAGATCGGAGGAAAGCCCTGCTTTATTTAAA 780

QY 781 GTTACCAATGACAAAGTGAATGGAATACTGTGATGCTCAGCCTGCTCAGCCAGGAC 840  
Db 781 GTTACCAATGACAAAGTGAATGGAATACTGTGATGCTCAGCCTGCTCAGCCAGGAC 840

QY 841 GTTCCCTACCCAGGAAAGCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC 900  
Db 841 GTTCCCTACCCAGGAAAGCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC 900

QY 901 TGTGAAAGAGCTGAGATAGCAGAGAGAGATCAAGAGAAATCTATGAGAGCTTTAAGAGC 960  
Db 901 TGTGAAAGAGCTGAGATAGCAGAGAGAGATCAAGAGAAATCTATGAGAGCTTTAAGAGC 960

QY 961 ACGGGGGCGACACCCATGGCAGGCGTCCCTCCAGTCCCTCGCTGCTGACCATCTCC 1020  
Db 961 ACGGGGGCGACACCCATGGCAGGCGTCCCTCCAGTCCCTCGCTGCTGACCATCTCC 1020

QY 1021 ATGCCCCCAGGCGCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080  
Db 1021 ATGCCCCCAGGCGCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080

QY 1081 GCCCACTGACCCGACATAAAAAACAGACATCTAAGGTGCTGTAGGGGACAGGACCTG 1140  
Db 1081 GCCCACTGACCCGACATAAAAAACAGACATCTAAGGTGCTGTAGGGGACAGGACCTG 1140

QY 1141 AAGAAAGAGAAATTTCTATGAGCAGAGCTTTAGGGTGCAGAGATATTCAAGTACAGCCAC 1200  
Db 1141 AAGAAAGAGAAATTTCTATGAGCAGAGCTTTAGGGTGCAGAGATATTCAAGTACAGCCAC 1200

QY 1201 TACAATGAAGAGATGAGATTTCCCACAATGATTTGCTTGTCTCAAGTAAAGCCAGTG 1260  
Db 1201 TACAATGAAGAGATGAGATTTCCCACAATGATTTGCTTGTCTCAAGTAAAGCCAGTG 1260

QY 1261 GATGTCACCTGTGCTCTAGAAATCCAAATAGTGAAGACTGTGTGCTTGTCTGATGGGTCC 1320  
Db 1261 GATGTCACCTGTGCTCTAGAAATCCAAATAGTGAAGACTGTGTGCTTGTCTGATGGGTCC 1320

QY 1321 TTTCCCTCTGGGAGTGAAGTCCCAATCTCTGGCTGGGGTGTTCAGAAAACAGGAAAAGGG 1380  
Db 1321 TTTCCCTCTGGGAGTGAAGTCCCAATCTCTGGCTGGGGTGTTCAGAAAACAGGAAAAGGG 1380

QY 1381 TCCCGCCAGCTCTTGGATGCCAAAGTCAAGTGAATGCCAACACTTTGTGCAACTCCCGC 1440  
Db 1381 TCCCGCCAGCTCTTGGATGCCAAAGTCAAGTGAATGCCAACACTTTGTGCAACTCCCGC 1440

QY 1441 CAACCTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
Db 1441 CAACCTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500

QY 1501 GGGGAGACACTCCAGGAGTGAATCTGGAGGCCCCCTGACCTGTGAGAGAGAGCGGACCC 1560  
Db 1501 GGGGAGACACTCCAGGAGTGAATCTGGAGGCCCCCTGACCTGTGAGAGAGAGCGGACCC 1560

QY 1561 TACTACGTCTATGGGATGAGTGGGGCTGGAGTGTGAGAAAGAGCCAGGGGTCTAC 1620  
Db 1561 TACTACGTCTATGGGATGAGTGGGGCTGGAGTGTGAGAAAGAGCCAGGGGTCTAC 1620

QY 1621 ACCCAAGTACCAATCTCTGAATGATCAAGCCACCATCAAAAGTGAAGTGGGTTC 1680  
Db 1621 ACCCAAGTACCAATCTCTGAATGATCAAGCCACCATCAAAAGTGAAGTGGGTTC 1680

QY	1681 TAA 1683     Db 1681 TAA 1683
RESULT 4	
BD177602	
LOCUS	BD177602 1683 bp DNA linear PAT 16-APR-2003
DEFINITION	Detection method with the use of factor VII activating protease mutant and specific antibody.
ACCESSION	BD177602
VERSION	BD177602.1 GI:30014864
KEYWORDS	JP 2002291486-A/1.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 1683) Roemisch,J., Stoehr,H.A., Feussner,A., Lang,W., Weimer,T., Becker,M., Nerlich,C. and Naumann,G.M.
TITLE	Detection method with the use of factor VII activating protease mutant and specific antibody
JOURNAL	Patent: JP 2002291486-A 1 08-OCT-2002; AVENTIS BEHRING GMBH
COMMENT	OS Homo sapiens (human) PN JP 2002291486-A/1 PD 08-OCT-2002 PF 25-JUL-2001 JP 2001224423 PR 26-JUL-2000 DE 10036641:4,10-OCT-2000 DE 10050040:4 PR 21-OCT-2000 DE 10052319:6,12-APR-2001 DE 10118706:8 PI JUERGEN ROEMISCH,HANS ARNOLD STOEHR,ANNETTE FEUSSNER,WIEGAND PI LANG, PI THOMAS WEIMER,MARGRET BECKER,CLAUDIA NERLICH,GUDRUN MUTH PI NAUMANN PC C12N15/09,C07K16/40,C12N9/50,C12Q1/68,G01N33/53,G01N33/533, PC GOIN33/534, PC GOIN33/535//C12P21/08,C12N15/00 CC Detection method with the use of factor VII activating CC protease mutant and CC specific antibody Location/Qualifiers FH Key Location/Qualifiers FT source 1..1683 FT /organism='Homo sapiens (human)'. Location/Qualifiers source 1..1683 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"
FEATURES	
source	
ORIGIN	
Query Match	99.8%; Score 1679.8; DB 6; Length 1683;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 1681; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
QY	1 ATGTTTCCCAGGATGTCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60
Db	1 ATGTTTCCCAGGATGTCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60
QY	61 GCCTGTGGGTTCCTCCCTGATGTCCTTTATTGGAAAGCCTGGACCAGACTGCACCCCTGAC 120
Db	61 GCCTGTGGGTTCCTCCCTGATGTCCTTTATTGGAAAGCCTGGACCAGACTGCACCCCTGAC 120
QY	121 CAGTATGATTACAGCTACGAGGATTATAATCAGAAGAGAACACCAGTAGCACATTACC 180
Db	121 CAGTATGATTACAGCTACGAGGATTATAATCAGAAGAGAACACCAGTAGCACATTACC 180
QY	181 CATGCTCAGAAATCCTGTA CTGGTACTACACTCAGAGCAACGATCATCATGCCAGCCCCAAC 240
Db	181 CATGCTCAGAAATCCTGTA CTGGTACTACACTCAGAGCAACGATCATCATGCCAGCCCCAAC 240
QY	241 CCCTGTGAACACGGTGGGAGCTGCTCGTGTCATGGAGACACCTTTCATGACGAGCTGCTG 300
Db	241 CCCTGTGAACACGGTGGGAGCTGCTCGTGTCATGGAGACACCTTTCATGACGAGCTGCTG 300





961 ACAGCGGGCAAGCACCATGGCAGGCGTCCCTCCAGCTCTCGCTCGCTCTGACCATCTCC 1020  
Db  
1057 ACAGCGGGCAAGCACCATGGCAGGCGTCCCTCCAGCTCTCGCTCGCTCTGACCATCTCC 1116  
Qy  
1021 ATGCCCCAGGCGCACCTTCTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCATGCT 1080  
Db  
1117 ATGCCCCAGGCGCACCTTCTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCATGCT 1176  
Qy  
1081 GCCCACTGCACCGACATAAATAACAGACATCTAAAGTGTGTAGGGACCAAGGACCTG 1140  
Db  
1177 GCCCACTGCACCGACATAAATAACAGACATCTAAAGTGTGTAGGGACCAAGGACCTG 1236  
Qy  
1141 AAGAAAGAGAAATTCATGAGCAGAGCTTTAGGGTGCAAGATATTCAGATGACAGCCAC 1200  
Db  
1237 AAGAAAGAGAAATTCATGAGCAGAGCTTTAGGGTGCAAGATATTCAGATGACAGCCAC 1296  
Qy  
1201 TACAATGAAGAGATGAGATTCGCCCAATGATATTCATGCTCAAGTTAAAGCCAGTG 1260  
Db  
1297 TACAATGAAGAGATGAGATTCGCCCAATGATATTCATGCTCAAGTTAAAGCCAGTG 1356  
Qy  
1261 GATGGTCACTGTGCTTAGAATCAAAATACGTGAAGACTGTGTGCTTGTGATGGGTCC 1320  
Db  
1357 GATGGTCACTGTGCTTAGAATCAAAATACGTGAAGACTGTGTGCTTGTGATGGGTCC 1416  
Qy  
1321 TTTCCCTCTGGAGTGAGTGCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1380  
Db  
1417 TTTCCCTCTGGAGTGAGTGCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1476  
Qy  
1381 TCCGCCAGCTCTCTGGATGCCAAAGTCAAGTGATTCGCCCAACTTTGTCAACTCCCGC 1440  
Db  
1477 TCCGCCAGCTCTCTGGATGCCAAAGTCAAGTGATTCGCCCAACTTTGTCAACTCCCGC 1536  
Qy  
1441 CAACCTATGACACATGATGATGACAGATGATCTGTGCGAGAAATCTTCAGAAACCT 1500  
Db  
1537 CAACCTATGACACATGATGATGACAGATGATCTGTGCGAGAAATCTTCAGAAACCT 1596  
Qy  
1501 GGGCAAGACACCTGCGAGGGTGAATCTGGAGGCCCCCTGACCTGTGAGAGGACGGCAC 1560  
Db  
1597 GGGCAAGACACCTGCGAGGGTGAATCTGGAGGCCCCCTGACCTGTGAGAGGACGGCAC 1656  
Qy  
1561 TACTACGTCTATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620  
Db  
1657 TACTACGTCTATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1716  
Qy  
1621 ACCCAAGTTACCAATTCCTGAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680  
Db  
1717 ACCCAAGTTACCAATTCCTGAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1776  
Qy  
1681 TAA 1683  
Db  
1777 TAA 1779

RESULT 6  
AX409021  
LOCUS AX409021 3008 bp DNA linear PAT 14-JUN-2002  
DEFINITION Sequence 1668 from Patent WO0229103.  
ACCESSION AX409021  
VERSION AX409021.1 GI:21441726  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE  
AUTHORS Alves, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.  
TITLE Gene expression profiles in liver cancer  
JOURNAL Patent: WO 0229103-A 1668 11-APR-2002;  
GENE LOGIC INC (US)  
FEATURES  
Location/Qualifiers  
1..3008  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"  
/note="EMBL/GenBank Accession No. D49742"  
Query Match 99.8%; Score 1679.8; DB 6; Length 3008;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
ORIGIN  
Qy 1 ATGTTTGCAGGATGTCTGATCTCCATGTTCTGCTGTTTAATGGCTCTGGTGGGAAAGACA 60  
Db 97 ATGTTTGCAGGATGTCTGATCTCCATGTTCTGCTGTTTAATGGCTCTGGTGGGAAAGACA 156  
Qy 61 GCCTGTGGGTTCCTCCATGATCTTTATTTGAAAGCTTGGACCCAGACTGACCCCTGAC 120  
Db 157 GCCTGTGGGTTCCTCCATGATCTTTATTTGAAAGCTTGGACCCAGACTGACCCCTGAC 216  
Qy 121 CAGTATGATTAACAGCTACGAGATTTAATCAGGAAGAGAACACAGTAGCACATTACC 180  
Db 217 CAGTATGATTAACAGCTACGAGATTTAATCAGGAAGAGAACACAGTAGCACATTACC 276  
Qy 181 CATGTGAGATTCCTGATCTGTCTACACTGAGGACCAAGCTGATCCATGCGAGCCCAAC 240  
Db 277 CATGTGAGATTCCTGATCTGTCTACACTGAGGACCAAGCTGATCCATGCGAGCCCAAC 336  
Qy 241 CCTGTGAACACGCTGGGAGCTGCTCCTCCATGGGAGCACTTCACATGCACTGCTGCTG 300  
Db 337 CCTGTGAACACGCTGGGAGCTGCTCCTCCATGGGAGCACTTCACATGCACTGCTGCTG 396  
Qy 301 GCTCTTTCTCTGGAATAAGTGTGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 360  
Db 397 GCTCTTTCTCTGGAATAAGTGTGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 456  
Qy 361 GSCCGGGGCAATGCTCATTTACCCAGAGTCTCCTCCTACTACCGCTGTGTCTGTAACAC 420  
Db 457 GSCCGGGGCAATGCTCATTTACCCAGAGTCTCCTCCTACTACCGCTGTGTCTGTAACAC 516  
Qy 421 CTTTACACAGTCTCCAGTGTCTCCAAAGTGTCTGTATGAGGCGCAAAACCCCTGCCAG 480  
Db 517 CTTTACACAGTCTCCAGTGTCTCCAAAGTGTCTGTATGAGGCGCAAAACCCCTGCCAG 576  
Qy 481 AATGGGCTACCTGCTCCCGGATAGCGGAGATCCAAAGTTCACTGTGCTGCTGCTGCCAG 540  
Db 577 AATGGGCTACCTGCTCCCGGATAGCGGAGATCCAAAGTTCACTGTGCTGCTGCCAG 636  
Qy 541 CAGTTCAAGGGGAAATTTCTGAAATAGTGTCTGATGATGCTATGTTGGGATGGCTAC 600  
Db 637 CAGTTCAAGGGGAAATTTCTGAAATAGTGTCTGATGATGCTATGTTGGGATGGCTAC 696  
Qy 601 TCTTACCGAGGAAATGATAGGACAGTCAACAGCATGGTGTGCTTTTACTGGAACTCC 660  
Db 697 TCTTACCGAGGAAATGATAGGACAGTCAACAGCATGGTGTGCTTTTACTGGAACTCC 756  
Qy 661 CACCTCCTCTTGCAGGAGAAATTAACCATGTTTATGGAGGATGCTGAAACCCATGGGATT 720  
Db 757 CACCTCCTCTTGCAGGAGAAATTAACCATGTTTATGGAGGATGCTGAAACCCATGGGATT 816  
Qy 721 GGGGAAACAATTTCTGAGAAACCCAGATGCGGAGAAAGCCCTGTGCTTTTATAA 780  
Db 817 GGGGAAACAATTTCTGAGAAACCCAGATGCGGAGAAAGCCCTGTGCTTTTATAA 876  
Qy 781 GTTACCAATGACAAAGTGAATGGGAATACTGTGATGCTCAGCTGCTCAGCCAGGAC 840  
Db 877 GTTACCAATGACAAAGTGAATGGGAATACTGTGATGCTCAGCTGCTCAGCCAGGAC 936  
Qy 841 GTTGCCTTCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGCGGGTTGACTCC 900  
Db 937 GTTGCCTTCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGCGGGTTGACTCC 996  
Qy 901 TGTGGAAGACTGAGATGACAGAGGAGAGATCAAGAGATCTATGAGGCTTTTAAGAGC 960  
Db 997 TGTGGAAGACTGAGATGACAGAGGAGAGATCAAGAGATCTATGAGGCTTTTAAGAGC 1056  
Qy 961 ACAGCGGGCAAGCACCCATGGCAGGCGTCCCTCCAGTCTCTGCTGCTCTGACCATCTCC 1020

1057	Db	ACGGCGGCAAGCACCCATGGCAGGCGTCCTCCAGTCTCTCGTGTGCTCTGACCAFTCTCC	1116
1021	Qy	ATGCCCAAGGGCCATTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT	1080
1117	Db	ATGCCCAAGGGCCATTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT	1176
1081	Qy	GCCCACTGCACCACACATAAAAACACAGACACTCTAAAGTGGTGCTAGGGACACGAGCACTG	1140
1177	Db	GCCCACTGCACCACACATAAAAACACAGACACTCTAAAGTGGTGCTAGGGACACGAGCACTG	1236
1141	Qy	AAGAAAGAGAATTTTCATGACGACAGAGTTTGGGTGCGAAGATATCTCAAGTACAGCCAC	1200
1237	Db	AAGAAAGAGAATTTTCATGACGACAGAGTTTGGGTGCGAAGATATCTCAAGTACAGCCAC	1296
1201	Qy	TACAATGAAAGAGATGAGATTTCGCCACAATGATATTTGCATTGCTCAAGTTAAAGCCAGTG	1260
1297	Db	TACAATGAAAGAGATGAGATTTCGCCACAATGATATTTGCATTGCTCAAGTTAAAGCCAGTG	1356
1261	Qy	GATGGTCACTGTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGGTCC	1320
1357	Db	GATGGTCACTGTGCTCTAGAAATCCAAATACGTGAAGACTGTGTGCTTGCCTGATGGGTCC	1416
1321	Qy	TTTCCCTCTGGGAGTGAGTGCCACATCTCTGGCTGGGGTGTTACAGAAACAGGAAAAGGG	1380
1417	Db	TTTCCCTCTGGGAGTGAGTGCCACATCTCTGGCTGGGGTGTTACAGAAACAGGAAAAGGG	1476
1381	Qy	TCCGCGAGCTCTGGATGCCAAAGTCAAGCTCAATTTGCCAACACTTTGTGCACTCCCGC	1440
1477	Db	TCCGCGAGCTCTGGATGCCAAAGTCAAGCTCAATTTGCCAACACTTTGTGCACTCCCGC	1536
1441	Qy	CAACTCTATGACCACATGATTGTATGACGTATGATCTGTGCAGGAAATCTTCAGAAACCT	1500
1537	Db	CAACTCTATGACCACATGATTGTATGACGTATGATCTGTGCAGGAAATCTTCAGAAACCT	1596
1501	Qy	GGGCAAGACACCTGCCAGGGTGACTCTGGAGGCCCTGACCTGTGAGAGGACGGCACCC	1560
1597	Db	GGGCAAGACACCTGCCAGGGTGACTCTGGAGGCCCTGACCTGTGAGAGGACGGCACCC	1656
1561	Qy	TACTAGCTATGGGATAGTGAAGCTGGGGCTGGAGTGTGAGAAGAGCGCAGGGTCTAC	1620
1657	Db	TACTAGCTATGGGATAGTGAAGCTGGGGCTGGAGTGTGGGAAGAGCGCAGGGTCTAC	1716
1621	Qy	ACCCAAGTTACCAAAATTCCTGAATGGATCAAGGCCACCATCAAAAGTGAAGTGGCTTC	1680
1717	Db	ACCCAAGTTACCAAAATTCCTGAATGGATCAAGGCCACCATCAAAAGTGAAGTGGCTTC	1776
1681	Qy	TAA	1683
1777	Db	TAA	1779

RESULT 7	HUMHGFAF	3008 bp	mRNA	linear	PRI 10-FEB-1999
LOCUS	HUMHGFAF				
DEFINITION	Human mRNA for HGF activator like protein, complete cds.				
ACCESSION	D49742				
VERSION	D49742.1	GI:736706			
KEYWORDS	HGF activator like protein; serin protease.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 3008)				
AUTHORS	Kitamura,N.				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 3008)				
AUTHORS	Kitamura,N.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-MAR-1995) Naomi Kitamura, Institute for Liver				
	Research, Kansai Medical University; Moriguchi, Osaka 570, Japan				
	(Tel:06-992-1001(ex.2530), Fax:06-994-6099)				
FEATURES	Location/Qualifiers				

```

source      1..3008
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /gb_xref="GI:134538"
            /translation="MPARMSDLHLVLLMALVGKTACGFSLMSLISLSLDPDWTPDOYDY
SYEDYNQEBENTSTTHAENPWTETDDQDFCNPCBHGGLVGHGFTFCSCSLAP
FSGNKQKQNTCKNCPGRGGCLITQSPYRVCVKHPYTGPSSQGVVCRPNPCQ
NGATCSRHRKTCFKACDPDQFKFGCEIGSDDCVYGDSYRGKNMTRNHQACLWY
NSHLLILNNRMFMDEATHGICGNFCNPDADEKPMWCIFIKVNDKVWEYCDVAS
SLTDVAVPESPTSESTKLPGPSOQCKTIAERKIIRIYGGFKTAGKEFWASLOSS
LPITSMVOHCFCGALIPVWLTAHQCTIDIKRLKVLVDQDLKEEFHFHSQSRV
EXLFKYSHNERDEIPHNDIALLKPFVDGHCALBSKYVKTCLPDGPSPPSGECHIS
GWGVETGKSQQLDKAVKYLANTLCNSRLQYDHMLDDSMICAGNLKPQGDTCCGD
SGLPLTECKDGYTVYGVISWLECGREGVITQTFLNLWKATIKSES GF"
polyA_site 2312
polyA_site 3008
ORIGIN
Query Match          99.8%; Score 1679.8; DB 9; Length 3008;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy   1  ATGTTTGCCAGGATGCTCATCTCCAATGTTCTGCTGCTTTAATGGCTCTGTGTGGAAAGACA 60
Db   97  ATGTTTGCCAGGATGCTCATCTCCAATGTTCTGCTGCTTTAATGGCTCTGTGTGGAAAGACA 156
Qy   61  GCCTGTGGGTTCCTCCTGATGCTCTTTATTGGAAAGCCTGGACCACACTGGACCCCTGCAC 120
Db   157 GCCTGTGGGTTCCTCCTGATGCTCTTTATTGGAAAGCCTGGACCACACTGGACCCCTGCAC 216
Qy   121 CAGTATGATTACAGCTACGAGGTTATATCAGGAAGAGAACAACAGTAGCACACTTACC 180
Db   217 CAGTATGATTACAGCTACGAGGATTATAATCAGGAAGAGAACAACAGTAGCACACTTACC 276
Qy   181 CATGCTGAGAATCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCCAAC 240
Db   277 CATGCTGAGAATCCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCCAAC 336
Qy   241 CCTGTGAAACAGGGTGGGATCGCTCTGTCATCGGAGACACTTACATGACGACTGCCTG 300
Db   337 CCTGTGAAACAGGGTGGGATCGCTCTGTCATCGGAGACACTTACATGACGACTGCCTG 396
Qy   301 GCTCCCTTCTCTCGGGAATAGTGTCAGAAAGTGC AAAATACGTCGCAAGGACCAACCCATGT 360
Db   397 GCTCCCTTCTCTCGGGAATAGTGTCAGAAAGTGC AAAATACGTCGCAAGGACCAACCCATGT 456
Qy   361 GGCGGGGCGCAATGCTCTATTACCGAGCTCTCCCTACTACCGCTGTGCTCTGAACAC 420
Db   457 GGCGGGGCGCAATGCTCTATTACCGAGCTCTCCCTACTACCGCTGTGCTCTGAACAC 516
Qy   421 CCTTACACAGGTCCCAGCTGCTCCCAAGTGGTTCCCTGTATGCAAGGCCAAAACCCCTGCCAG 480
Db   517 CCTTACACAGGTCCCAGCTGCTCCCAAGTGGTTCCCTGTATGCAAGGCCAAAACCCCTGCCAG 576
Qy   481 AATGGGGCTACTGCTCTCCGGCATACCGGAGATCCAAGTTTACCTGTGCTGCTGCCGAC 540
Db   577 AATGGGGCTACTGCTCTCCGGCATACCGGAGATCCAAGTTTACCTGTGCTGCTGCCGAC 636
Qy   541 CAGTTCAAGGGGAAATTCCTGTGAAATAGGTTCTGATGACTGCTATGTGCGGATGGCTAC 600
Db   637 CAGTTCAAGGGGAAATTCCTGTGAAATAGGTTCTGATGACTGCTATGTGCGGATGGCTAC 696
Qy   601 TCTTACCGAGGAAATGAAATAGGACAGTCAAACGACGATGGCTCTTACTGGACTCC 660
Db   697 TCTTACCGAGGAAATGAAATAGGACAGTCAAACGACGATGGCTCTTACTGGACTCC 756
Qy   661 CACCTCCTCTTGCGAGGAAATTACACATGTTTTATGGAGGATGCTGAAACCCCATGGGATT 720

```



681	Db	TCTTACCGAGGGAAAATGAATAGGACAGTCACACGATGCTGGTGCCCTTTACTCGAACTCC	7140
661	Qy	CACCTCCTCTTTCAGGAGAAATTACAAATGTTTATGAGGATGCTGAAACCCATGGGATT	720
741	Db	CACCTCCTCTTTCAGGAGAAATTACAAATGTTTATGAGGATGCTGAAACCCATGGGATT	800
721	Qy	GGGGAACACAAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTGGTCTTATATAA	780
801	Db	GGGGAACACAAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTGGTCTTATATAA	860
781	Qy	GTTTACCAATGACAAGTGGAAATGGGAATACTGTGATGTCTCAGCCTGCTCAGCCAGGAC	840
861	Db	GTTTACCAATGACAAGTGGAAATGGGAATACTGTGATGTCTCAGCCTGCTCAGCCAGGAC	920
841	Qy	GTTTGCCTTACCCAGAGAAACCCCATCTGAGCCATCAACCAAGCTTCGCGGGTTGACTCC	900
921	Db	GTTTGCCTTACCCAGAGAAACCCCATCTGAGCCATCAACCAAGCTTCGCGGGTTGACTCC	980
901	Qy	TGTGGAAAGACTCAGATAGCAGAGAGGAAGATCAAGAGAAATCTATGAGAGCTTTAAGAGC	960
981	Db	TGTGGAAAGACTCAGATAGCAGAGAGGAAGATCAAGAGAAATCTATGAGAGCTTTAAGAGC	1040
961	Qy	ACGGCGGGCAAGCAACCATGGCAGGGCTCCCTCCAGTCTTCGTGCTCTGACCATCTCC	1020
1041	Db	ACGGCGGGCAAGCAACCATGGCAGGGCTCCCTCCAGTCTTCGTGCTCTGACCATCTCC	1100
1021	Qy	ATGCCCAAGGCGACATTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT	1080
1101	Db	ATGCCCAAGGCGACATTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT	1160
1081	Qy	GCCTACTGCAACGACATATAAAACCCAGACATCTAAAGGTGGTGTAGGGGACCAAGGACCTG	1140
1161	Db	GCCTACTGCAACGACATATAAAACCCAGACATCTAAAGGTGGTGTAGGGGACCAAGGACCTG	1220
1141	Qy	AAGAAAGAGAAATTTCAATGACGAGAGCTTTAGGGTGCAGAGATATTTCAAGTACAGCCAC	1200
1221	Db	AAGAAAGAGAAATTTCAATGACGAGAGCTTTAGGGTGCAGAGATATTTCAAGTACAGCCAC	1280
1201	Qy	TACAATGAAAGAGATGAGATTCCTCCCAACATGATATTGCATTTGCTCAAGTTAAAGCCAGTG	1260
1281	Db	TACRAATGAAGAGATGAGATTCCTCCCAACATGATATTGCATTTGCTCAAGTTAAAGCCAGTG	1340
1261	Qy	GATGGTCATCTGTGCTTAGAAATCCAAATAGCTGAAGACTGTGTGCTTGCCTGATGGGTCC	1320
1341	Db	GATGGTCATCTGTGCTTAGAAATCCAAATAGCTGAAGACTGTGTGCTTGCCTGATGGGTCC	1400
1321	Qy	TTTTCCCTCTCGGAGTGTGCCACATCTCTGGCTGGGGTGTTCACAGAAAACAGGAAAAAGG	1380
1401	Db	TTTTCCCTCTCGGAGTGTGCCACATCTCTGGCTGGGGTGTTCACAGAAAACAGGAAAAAGG	1460
1381	Qy	TCGCGCGAGCTCCTGGATGCCAAAGTCAAGCTGATTTGCCAAACACTTTTGTGCAATCCCGC	1440
1461	Db	TCGCGCGAGCTCCTGGATGCCAAAGTCAAGCTGATTTGCCAAACACTTTTGTGCAATCCCGC	1520
1441	Qy	CAACTCTATGACCACATGATTTGATGACAGTATGATCTGTGACGAAATCTTTCAGAAACCT	1500
1521	Db	CAACTCTATGACCACATGATTTGATGACAGTATGATCTGTGACGAAATCTTTCAGAAACCT	1580
1501	Qy	GGGCAAGACACCTGCCAGGGTGAATCTTGGAGGCCCTTGCACCTGTGAGAGGACGGCAC	1560
1581	Db	GGGCAAGACACCTGCCAGGGTGAATCTTGGAGGCCCTTGCACCTGTGAGAGGACGGCAC	1640
1561	Qy	TACTACGTCTATGGGATAGTGTGGGCTTGGAGTGTGAGAGAGCCAGGGGTCTAC	1620
1641	Db	TACTACGTCTATGGGATAGTGTGGGCTTGGAGTGTGAGAGAGCCAGGGGTCTAC	1700
1621	Qy	ACCCAAGTTACCAAAATTCCTGAATTTGGATCAAAAGCCACCATCAAAAGTGAAGTGGGTTC	1680
1701	Db	ACCCAAGTTACCAAAATTCCTGAATTTGGATCAAAAGCCACCATCAAAAGTGAAGTGGGTTC	1760
1681	Qy	TAA 1683	
1761	Db	TAA 1763	

RESULT 9  
AK128915  
LOCUS  
DEFINITION  
Mus musculus cDNA fis, clone TRACH2024735, moderately similar to  
Homo sapiens hyaluronan binding protein 2 (HABP2).  
ACCESSION  
AK128915  
VERSION  
AK128915.1 GI:34536572  
KEYWORDS  
oligo capping; fis (full insert sequence).  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Chordata; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1  
AUTHORS  
Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,  
Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K.,  
Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M.,  
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A.,  
Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K.,  
Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y.,  
Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K.,  
Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,  
Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and Isogai, T.  
TITLE  
NEDO cDNA sequencing project  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 2190)  
AUTHORS  
Isogai, T. and Yamamoto, J.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
COMMENT  
NEDO cDNA sequencing project supported by Ministry of Economy,  
Trade and Industry of Japan; cDNA full insert sequencing; Research  
Association for Biotechnology (RAB); cDNA library construction;  
Helix Research Institute (HRI) (supported by Japan Key Technology  
Center etc.); 5' & 3'-end one pass sequencing; RAB, HRI, and  
Biotechnology Center, National Institute of Technology and  
Evaluation; Clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.  
FEATURES  
source  
1..2190  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="TRACH2024735"  
/note="cloning vector: pME18SFL3"  
ORIGIN  
Query Match 67.8%; Score 1141; DB 10; Length 2190;  
Best Local Similarity 80.3%; Pred. NO. 0;  
Matches 1352; Conservative 0; Mismatches 325; Indels 6; Gaps 1;  
QY 1 ATGTTTGCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60  
DB 62 ATATTTGTCCAGATGTTGGTGTTCCTGCTGCTGCTAATCGCCCTGCTGGGAAGTCA 121  
QY 61 GCTGTGGGTCTCCCTGATGCTTTATTGGAAGCCTGGACCCAGACTGGACCCCTGAC 120  
DB 122 GTCAATTGGGCTCTCACTGATGCTTCAATGCGCCCCAGACCCAGATGGACCCCGAT 181  
QY 121 CAGTATGATTACAGCTACGAGGATTATATACAGGAAGAACACACAGTAGCACCTTACC 180  
DB 182 GACTATTACTACAGCTATGAGCAGTCCAGCCAGACGAGACCCAGTGTCAAGCAGACC 241  
QY 181 CATGCTGAGATCTGACTGGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240  
DB 242 ACCCTTGAGAACCCGACCTGGTACT-----ATGAAGACGATGATCCATGCCAGTCCAAC 295  
QY 241 CCCTGTGAACACGGTGGGACTGCCTCGTCCATGGGAGCACCTTCACATGACGCTGCCTG 300  
DB 296 CCCTGTGAACACGGGAGTGTATCATCAGGGGATACCTTCAGTTGCAAGCTGCCCA 355

QY 301 GCTCTTCTCTCGGGAATAAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAAACCATGT 360  
DB 356 GCGCCCTTCTCGGGGAGCCGGTGCCAGACTGCACAAAACAAGTGCAGGACAAACCCATGT 415  
QY 361 GGCGGGGGCAATGTCTCATTATCCAGAGACTCCTCCCTACTACCGCTGTGTCTGTAAACAC 420  
DB 416 GTCCATGGTGAATTGCGCTCATTTACCCAGAGACACCCCTACTACCGCTGTGCTGCAATAC 475  
QY 421 CTTTACACAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGACCAACCCCTGCCAG 480  
DB 476 CTTTACACGGGACAGACTGCTCCAAAGTGTTCGGGCAATGAGGCAACCCCTGCCAG 535  
QY 481 AATGGGGTACCTGCTCCCGGATAGCGAGATCCAAAGTTCACCTGTGCTGCTGCCGAC 540  
DB 536 AATGGGGAGTCTGTTCCCGACACAGCGAGATCCAGGTTTACCTGTGCTGCTGCCAGAC 595  
QY 541 CAGTCAAGGGGAAATCTGTGAATAGTGTCTGTAGTCTGTATGTTGGCGATGGCTAC 600  
DB 596 CAGTAAAGGGGAAATCTGTGAATAGTGTCCGAGACTGTGTATGTCGGTGTATGGCTAC 655  
QY 601 TCTTACCGAGGGAAATGAATAGGACAGTCAACAGCATGCGTGCCTTTACTTGGAACTCC 660  
DB 656 TCTTACCGAGGCAAGTGAAGTAAAGTCAACAGACCCATGCCCTTTACTTGGAACTCC 715  
QY 661 CAGTCTCTTGTGAGGAGAAATTAACAATGTTTATGAGAGATGCTGAAACCCATGGGAT 720  
DB 716 CAGCTCTCTTGTGAGGAGACTTATAACATGTTTATGAGGATGTCAGAGACCCACGGGATC 775  
QY 721 GGGGAACAACAATTTCTGCAGAAACCCAGATCGGAGCAAAAGCCCTGGTGTCTTATATAA 780  
DB 776 GCAGAGCAACAATCTTCTGCAGAAACCCAGATGGAGACCAACAACCTGGTGTCTCGTCAAG 835  
QY 781 GTTACCAATGACAAGGTGAATAATGGGAATCTGTGATGTCTAGCTGTCTCAGCCACAGGAC 840  
DB 836 GTGAACAGTGAAGGTGAATAATGGGAATCTGTGATGTCTACAGTCTGTCCAGTGCCTGAC 895  
QY 841 GTTGCCTTACCAGAGAAAGCCCACTGAGGACATCAACCAAGCTTCCGGGGTTTGACTCC 900  
DB 896 ACCCTTAACCACTGGAAAGCCTTCTGGAGCTGTGATGGAGTGCAGGTTCCAGTCTCC 955  
QY 901 TGTGGAAGAAGCTGAGATAGCAGAGAGGAAGATCAAGAGAATCTATGAGAGCTTTAAGAGC 960  
DB 956 TGCAGGAAGACGGAGGTAGCTGAACACGAGTCAAGCGTATCTACGGGGCTTTAAGAGC 1015  
QY 961 ACGGGGGGAAGCACCCATCGGAGGCGTCCCTCAGTCTCGTCTGCTGCTGACCTCTCC 1020  
DB 1016 ACAGCAGGCAAGCACCCCGTGGCAGGTGTCCCTGACAGCTCACTGCGGTGACCACTCC 1075  
QY 1021 ATGCCCCAGGGCCCACTTCTGTGGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080  
DB 1076 ATGCCCCAAGGCCACTTCTGTGGGGGCGCCCTGATCCACCCCTGCTGGGTGCTCACTGCA 1135  
QY 1081 GCGCACTGCAACGACATAAACAACAGACATCTAAAGGTGGTGTAGGGACACAGGACCTG 1140  
DB 1136 GCGCACTGTACCGACATAAACACCAAGCATCTAAAGTGTGTACTAGGGGATCAGGACCTG 1195  
QY 1141 AAGAAAGAAGAAATTTCAATGAGCAGAGCTTTAGGGTGCAGAAAGATATTCAAGTACAGCCAC 1200  
DB 1196 AAGNAGACAGATCCCATGACAGACCTTCAGGGTGGAAAAAATACTGAATACAGTACAG 1255  
QY 1201 TACAATGAAGAGATGAGATTCGCCCAATGATATTGCTATGCTCAAGTTAAAGCCAGTG 1260  
DB 1256 TATAATGAAGAGATGAGATTCGCCCAATGATGCTTTGCTCAAGTTAAAGCCAGTG 1315  
QY 1261 GATGGTCACTGTGCTCTAGAAATCCAAATACGTCAAGACTGTGTGCTTGTGCTGATGGTCC 1320  
DB 1316 GTGGTCACTGTGCTCTGGAAATCCAGATATGAGACTGTATGTTGGCCAGGACCCC 1375  
QY 1321 TTTCCCTCTCGGAGTGAAGTCCCATCTCTGGTGGGGTGTACAGAAACAGAAAGGG 1380  
DB 1376 TTTCCCTCTGAACTGAGTGCACATCTCTGGTGGGGTGTACAGAAACAGGGAAGGG 1435  
QY 1381 TCCGCCAGCTCTCTGGATGCCAAGTCAAGCTGATTCGCAACACTTTGTGCACTCCCGC 1440

1436	TCCGCGCAGCTCTGGGATGCTAAAGTCACAGCTATCGCTMACCCCTTTGTGCACATCCCGA	1495
Db		
1441	CAACTCTATGACCACTGATGTGATGATGATCTGTGACGAGAAATCTTTCAGAAACCT	1500
Qy		
1496	CAACTCTATGACCACTGATGTGATGATGATGTGTGGGGGAACTTTCAGAAAGCCC	1555
Db		
1501	GGGCAAGACACTGCCAGGGTGACTCTGGAGGCCCTCTGACCTGTGAGAGGACGCGACC	1560
Qy		
1556	GGATCAGACACTGCCAGGGTGACTCGGGGGGCCCTCTAACTGTGAGAGGATGGAACT	1615
Db		
1561	TACTACGTCTATGGGATGATGACCTGGGGCCCTGGAGTGTGAGAGAGCCAGGGGTCTAC	1620
Qy		
1616	TACTACGTCTACGGGATGTAACTGGGGCCAGGAATGTGGGAGAGCCAGAGATCTAC	1675
Db		
1621	ACCCAGTTACCAATTCCTGAATTCGATCAAGCCACCATCAAAAGTGAAGTGGCTTC	1680
Qy		
1676	ACTCAAGTCACCAAGTTCCTGAATTCGATTAAGACCAACCATGACAGGAGGCTGGCCTC	1735
Db		
1681	TAA 1683	
Qy		
1736	TGA 1738	
Db		

RESULT 10	
BC031775	BC031775
LOCUS	2119 bp mRNA linear ROD 04-OCT-2003
DEFINITION	Mus musculus hyaluronic acid binding protein 2, mRNA (cDNA clone MGC:28705 IMAGE:424257), complete cds.

ACCESSION BC031775  
VERSION BC031775.1 GI:21594554  
KEYWORDS MGC.

ORGANISM  
Mus musculus  
Mus musculus (house mouse);  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 2119)  
REFERENCE

Klausner, R.D., Collins, F.S., Wagner, L., Shermen, C.M., Schuler, G.D., Alschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Shat, N.K., Hopkins, R.F., Jordan, H., Moore, T.L., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Journal of Molecular Evolution

REFERENCE  
 PUBMED  
 12477932  
 2 (bases 1 to 2119)

**TITLE**  
Direct Submission  
Submitted (06-JUN-2002) National Institutes of Health, Mammalian

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

COMMENT

Contact: MGC help desk  
Email: cgapbs-r@mail.nih.gov  
nicole.precourt@tuffnau.com

CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center.  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: amgb@bcm.tmc.edu  
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounsgood, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Naravati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK part: 37 Row: n Column: 14  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA g: 22122612.

FEATURES  
source

```
1. 2119
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone=MGC:28705 IMAGE:4243577
/tissue_type="kidney, normal, 5 month old male mouse."
/clone_lib="NCI CGAP Kid14"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
```

gene

```

/gene="Hapb2"
/note="synonyms: MGC38705, HAPB, PHBP, HGFAI"
/db_xref="locusid:226243"
/db_xref="MGI:1196378"
59.1612

```

## CDS

```

/codon_start=1
/product="hyaluronic acid binding protein 2"
/protein_id="AAH31775.1"
/db_xref="GI:21594555"
/db_xref="LocusID:226243"

```

```

/translation="MIVFRVLLIIALVKGSVTLGLSMFAEPDPPDCQSNPCPEHGGS
CIIRGDTFSPCAFSGRSRQTAQKNDKPNVGGDCLLITQKHPYCRACKPYVTGP
CDSKYLPAACRNQGVCSHRRRSRTACQPKGCEIGDPDCCVGGDYSYR
GKSVTAQNQLVYNHLLQSTYNNMFEDATGIAEHNFCRNPDDHKKPQFVKV
NSEKWKVCDVYVCPDTPNPSLVEMLPEFSGESCETAEVAHAKVIRLYGFGK
STAGKHQVQISLTSLPTSMPOGHFCGAGLIPCWLLTAACHTINTKHLXWLCD
QDLKTESHEPTVREYKILKVSQYNERDIPNDITALLKLVKGHGHALESVRVVKV
LPSDPFPGTECHISGWVGTTEGSRQQLDAKKLIANPLCNSRQLYDHTIDDSMLC
AGNLQKQSDTCQSGDGGPLTCEKQDGTYYVIGVSWGQBGKPKPVYTKELNWLK
TTHREAGI."

```

## misc feature

```

/region="KR; Region: Kringle domain"
/db_xref="CDD:smart00130"

```

misc_feature	866	.1579
--------------	-----	-------

```

/note="Tryp_Spc; Region: Trypsin-like serine protease"
/db_xref="CDD:smart00020"

```

## ORIGIN

```

ch 62.2%; Score 1047; DB 10; Length 2119;
l Similarity 82.2%; Pred. No. 7,7e-304;
203; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

```

221 CTGATCCATGSCAGCCCCAACCCCTGTGAACACGGTGGGGACTGCCTCGTCCATGGGAGCA 280

50 CAGATCCATGCCAGTCCAAACCCCTGTGAACACGGCGGGACTGTATCATCAGAGGGGATA 209

81 CCTTCACATGCAGCTGCCTGGCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGC AAAATA 340

10 CCTTCAGTTGCAGCTGCCAGGCCCTTCTCGGGAGCGCGTGCCAGACTGCACAAAACA 269

441 CGTGCAAGGCAACACCCATGTGGCCGGGGCCAAATGCTCATATACCCAGAGTCTCTCCCTACT 400

770 AGTGCAAGGACAACCCCATGTGTCCATGGTGTATGGCCCTCATTACCCAGAAACACCCCCIACCT 329  
01 ACCCGCTGTCTGTATAACACACCCCTTACACAGGTCCCAGTGTCTCCAACTGGTTCCTGTAT 460

330 ACCGGTGTGCTGCAATAACCTTACACGGGACAGACTGTCTCCAAGTGCTTCGGGCAT 389

.....



1541 CCTGTGAGAGACGGACCTACTACTGCTATGAGATGAGCTGGGCTGGAGTGTG 1600  
 1470 CCTGTGAGAGATGGAATCTACTACTGCTACTGGAATTTGTAAGCTGGGCGCCAGGAATGTG 1529  
 1601 AGAAGAGGCGGAGGTCTACACCCCAAGTTACCAAAATTCCTGAATTTGATCAAGACCA 1660  
 1530 GGAAGAGGCGGAGGTCTACACTCAAGTCACCAAGTCTCTGAATTTGATCAAGACCA 1589  
 1661 TCAAAAGTGAAGTGGCTTCTAA 1683  
 1590 TGCACAGGAGGCTGGCCTCTGA 1612

RESULT 11  
 AL390197  
 LOCUS  
 DEFINITION  
 Human DNA sequence from clone RP11-190F19 on chromosome 10,  
 complete sequence.  
 AL390197  
 VERSION  
 AL390197.19 GI:15387800  
 KEYWORDS  
 HTG.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Tracey, A.  
 REFERENCE  
 1 (bases 1 to 188349)  
 AUTHORS  
 Direct Submission  
 TITLE  
 Submitted (29-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 JOURNAL  
 requests: clonerequest@sanger.ac.uk  
 COMMENT  
 On Aug 31, 2001 this sequence version replaced gi:14596344.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em., EMBL; Sw.,  
 SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 10, constructed by the Sanger Centre Chromosome 10  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-190F19 is from the library RPCI-11.1 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACE3.6  
 This sequence is the entire insert of clone RP11-190F19 The true  
 right end of clone RP11-481H12 is at 111825 in this sequence.

FEATURES  
 Location/Qualifiers  
 1..188349  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-190F19"  
 /clone\_lib="RPCI-11.1"  
 22142..22187  
 /notes="Sequence from overlapping clone RP11-481H12  
 (AL133482). Assembly confirmed by restriction digest."

misc\_feature  
 ORIGIN

461 GCAGGCCAAACCCCTGCCAGAAATGGGCTACCTGCTCCCGGCATAAAGCGAGATCCCAAGT 520  
 390 GCAGGCCAAACCCCTGCCAGAAATGGGAGTCTGTTCGCCGACACAGACGGAGATCCAGGT 449  
 521 TCACCTGTGCTGTCCCGACAGTTCAGGGGAAATCTCTGGAATATAGTTCGTGATCACT 580  
 450 TTACCTGTGCTGTCCCGACAGTATAGGGGAAATCTCTGGAATATAGTTCGGACGACT 509  
 581 GCTATGTTGGGAGTGGCTTACTCTTACCAGAGGAAATGAATAGGACAGTCAACACAGCATG 640  
 510 GTTATGTCGGTGTGCTTACTCTTACCAGAGGAAATGAATAGGACAGTCAACACAGCATG 569  
 641 CGTGCCTTTACTGGAACCTCCACCTCTCTTTCAGAGGAGAAATTAACAATGTTTATGAGG 700  
 570 CATGCTTTTACTGGAACCTCCACCTCTCTTTCAGAGGAGAAATTAACAATGTTTATGAGG 629  
 701 ATGCTGAAACCCATCGGATTTGGGAACACAATTTCTGCAGAAACCCAGATCGGACGAAA 760  
 630 ATGCAGAGACCCACGGATTCGAGGACACACTCTCTGCAGAAACCCAGATCGGACGACA 689  
 761 AGCCCTGTGCTTTTAAAGTTACCAATGCAAGGTGAATGGAATGGAATGATGTCT 820  
 690 AACCTGTGTTTCTCAAGGTGAACAGTGAAGGTGAATGGAATGGAATGATGTCTCA 749  
 821 CAGCTGTCTCAGCCAGGAGCTTCTACCCAGAGGAAAGCCCACTGAGCCATCAACCA 880  
 750 CAGTCTGTCTCAGCTGCTGACACCCCTAACCCAGTGAAGGCTTCTGGAGCTGTGATGG 809  
 881 AGCTTCCGGGTTTGAATCTCTGTGAAAGACTGATAGCAGAGGAAAGATCAAGAGAA 940  
 810 AGCTGCCAGGTTTGAATCTCTGTGAAAGACTGATAGCAGAGGAAAGATCAAGAGAA 869  
 941 TCTATGAGGCTTTAAGACAGCGGCGGACACCCATGCGAGGAAAGATCAAGAGAA 1000  
 870 TCTACGGGGCTTTAAGACAGCGGCGGACACCCATGCGAGGAAAGATCAAGAGAA 929  
 1001 CGCTGCTCTGACCTCTCTCATCCCGGAGGCACTTCTGTGTGGGGCGCTGATCCACC 1050  
 930 CACTGCGGTGACCACTCTCATCCCGGAGGCACTTCTGTGTGGGGCGGCTGATCCACC 989  
 1061 CTTGCTGGTGTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1120  
 990 CTTGCTGGTGTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1049  
 1121 TGTAGGGGACAGGACCTTGAAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGCAGA 1180  
 1050 TACTAGGGGATCAGGACCTTGAAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGCAGA 1109  
 1181 AGATATTCAGTACAGCCACTCAATGAAAGAGATGAGATCCCAATGATATGTCAT 1240  
 1110 AAATCTGAAGTACAGTCAAGTATATGAAAGAGATGAGATCCCAATGATATGTCAT 1169  
 1241 TGTCTAAGTTAAGCCAGTGGATGGTCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1300  
 1170 TGTCTAAGTTAAGCCAGTGGTGGTCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1229  
 1301 TGTGCTTGTGCTGATGGGTCTTTTCCCTCTGCGAGTGAAGTGCACATCTCTGCTGGGTG 1360  
 1230 TATGTTTGGCCAGCGACCCCTTCCCTCTGGAAGTCAAGTGCACATCTCTGCTGGGTG 1289  
 1361 TTACAGAAACAGAAAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1420  
 1290 TTACAGAAACAGGGGAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1349  
 1421 ACATTTGTGCAACTCCCGCAACTCTATGACCAATGATGATGATGATGATGATGATGATG 1480  
 1350 ACCCTTTGTGCAACTCCCGCAACTCTATGACCAATGATGATGATGATGATGATGATG 1409  
 1481 CAGAAATCTTCAAGAACTGGGCAAGACCTGCGAGGTCACCTCTGAGGAGCCCTCTGA 1540  
 1410 CGGGGAACCTTTCAGAAAGCCCGGATCAGACCTGCGAGGTCACCTCTGAGGAGCCCTCTAA 1469



Db	57685	GAATAGCACAAATTTATCTTTCTTGTTGCCACAGACGTTGCTTACCACAGAGAAAGCCCC	57626
QY	865	ACTGAGCCATCAACCAAGCTTCGGGGTTTGACTCTGTGGAAGACTGAGATAGCAGAG	924
Db	57625	ACTGAGCCATCAACCAAGCTTCGGGGTTTGACTCTGTGGAAGACTGAGATAGCAGAG	57566
QY	925	AGGAAGATCAAGAGAAATCTATGGAGGCTTTTAAGAGCAGCGGGGCAAGCACCCTATGGCAG	984
Db	57565	AGGAAGATCAAGAGAAATCTATGGAGGCTTTTAAGAGCAGCGGGGCAAGCACCCTATGGCAG	57506
QY	985	GGGTCCCTCAAGTCTCTGGTCTGTAGCAATCTCCATGCCAGGCGCACTTCTGTGGT	1044
Db	57505	GGGTCCCTCAAGTCTCTGGTCTGTAGCAATCTCCATGCCAGGCGCACTTCTGTGGT	57446
QY	1045	GGGGGCTGATCACCCCTGCTGGGTGTCTCACTGTCGCCACTGCACCGA	1094
Db	57445	GGGGCACTGATCACCCCTGCTGGGTGTCTCACTGTCGCCACTGCACCGA	57396
RESULT 13			
AX305372		397 bp	DNA
LOCUS			linear
DEFINITION			Sequence 123 from Patent WO0186188.
ACCESSION			AX305372
VERSION			AX305372.1
KEYWORDS			GI:117644923
SOURCE			Mus musculus (house mouse)
ORGANISM			Mus musculus
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS			Ishikawa, K., Asai, S., Takahashi, Y., Nagata, T. and Ishii, Y.
TITLE			Method for examining ischemic conditions
JOURNAL			Patent: WO 0186188-A 123 22-NOV-2001; School Juridical Person Nihon University (JP)
FEATURES			Location/Qualifiers
source			1..397
			/organism="Mus musculus"
			/mol_type="unassigned DNA"
			/db_xref="taxon:10090"
ORIGIN			
Query Match			14.3%; Score 241; DB 6; Length 397;
Best Local Similarity			84.4%; Pred. No. 8.7e-61;
Matches			271, Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY	1363	ACAGAAACAGAAAAAGGTCCCGCAGCTCTGGATGCCAAGCTCAAGCTGATGGCCAAC	1422
Db	13	ACAGAAACAGGGGAAGGTCCCGCAGCTCTGGATGCCAAGCTCAAGCTGATGGCCAAC	72
QY	1423	ACTTGTGTCAACTCCCGCAACTCTATGACCACTGATGACAGTATGATCTGTGCA	1482
Db	73	CTTTGTGTCAACTCCCGCAACTCTATGACCACTGATGACAGTATGATTTGTGG	132
QY	1483	GGAAATCTTGTGAACTCTGGCAGACACTCTGCCAGGCTGACCTCGAGGCCCTTGACC	1542
Db	133	GGGAACCTTTCAGAAAGCCGGATCAGACACTCTGCCAGGTGACCTCGGGGGCCCTTAACC	192
QY	1543	TGTGAGAAGGACCGCACCTACTACGCTTATGGATAGTGAAGTGGGCGCTGGAGTGTGAG	1602
Db	193	TGTGAGAAGGATGAACTTACTACGCTTACGGGATGTTAAAGTGGGCGCCAGGAATGTGGG	252
QY	1603	AAGAGGCCAGGGTCTACACCCAGTTTACCAAACTCTGTAATTTGGATCAAGCCACCATC	1662
Db	253	AAGAGGCCAGAGGTCTACATCAAGTCACCAAGTTCTCTGAAATTTGGATTAAGACCAACCATG	312
QY	1663	AAAAGTGAAGCTGGCTTTCTAA	1683
Db	313	CACAGGAGGCTGGCTCTGA	333

AC115771

LOCUS

DEFINITION

AC115771

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC115771 204214 bp DNA linear HTG 18-JUN-2003  
Mus musculus clone RP23-116D4, WORKING DRAFT SEQUENCE, 11 ordered  
pieces.

AC115771 GI:31880182

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 204214)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus, clone RP23-116D4

Unpublished

2 (bases 1 to 204214)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barina,N., Baslet,V., Bloom,T., Boguski,M.,

Bouckgeater,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

Chazaro,B., Choepel,Y., Collings,S., Collins,S., Collamore,A.,

Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,

Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kelle,C., Lacroque,K., Lamazares,R.,

Lander,E., Leach,K., Levine,R., Lindblad-Toh,K., Liu,G.,

MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,

McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,

Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,

Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,

Raymond,C., Retta,R., Rieback,M., Riley,K., Rise,C., Rogov,P.,

Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,

Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 204214)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,

Anderson,M., Arachchi,H.M., Barina,N., Bastien,V., Bloom,T.,

Boguski,M., Bouckgeater,B., Camarata,J., Chang,J., Choepel,Y.,

Collamore,A., Cooke,P., Corum,B., DeArelano,K.,

Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,

Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,

Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,

Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kelle,C., Lander,E., Levine,R.,

Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,

AC115771

LOCUS

DEFINITION

AC115771

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC115771 204214 bp DNA linear HTG 18-JUN-2003  
Mus musculus clone RP23-116D4, WORKING DRAFT SEQUENCE, 11 ordered  
pieces.

AC115771 GI:31880182

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 204214)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus, clone RP23-116D4

Unpublished

2 (bases 1 to 204214)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barina,N., Baslet,V., Bloom,T., Boguski,M.,

Bouckgeater,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

Chazaro,B., Choepel,Y., Collings,S., Collins,S., Collamore,A.,

Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,

Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kelle,C., Lacroque,K., Lamazares,R.,

Lander,E., Leach,K., Levine,R., Lindblad-Toh,K., Liu,G.,

MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,

McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,

Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,

Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,

Raymond,C., Retta,R., Rieback,M., Riley,K., Rise,C., Rogov,P.,

Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,

Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 204214)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,

Anderson,M., Arachchi,H.M., Barina,N., Bastien,V., Bloom,T.,

Boguski,M., Bouckgeater,B., Camarata,J., Chang,J., Choepel,Y.,

Collamore,A., Cooke,P., Corum,B., DeArelano,K.,

Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,

Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,

Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,

Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kelle,C., Lander,E., Levine,R.,

Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,

AC115771 204214 bp DNA linear HTG 18-JUN-2003  
Mus musculus clone RP23-116D4, WORKING DRAFT SEQUENCE, 11 ordered  
pieces.

AC115771 GI:31880182

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 204214)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus, clone RP23-116D4

Unpublished

2 (bases 1 to 204214)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barina,N., Baslet,V., Bloom,T., Boguski,M.,

Bouckgeater,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

Chazaro,B., Choepel,Y., Collings,S., Collins,S., Collamore,A.,

Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,

Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kelle,C., Lacroque,K., Lamazares,R.,

Lander,E., Leach,K., Levine,R., Lindblad-Toh,K., Liu,G.,

MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,

McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,

Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,

Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,

Raymond,C., Retta,R., Rieback,M., Riley,K., Rise,C., Rogov,P.,

Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,

Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 204214)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,

Anderson,M., Arachchi,H.M., Barina,N., Bastien,V., Bloom,T.,

Boguski,M., Bouckgeater,B., Camarata,J., Chang,J., Choepel,Y.,

Collamore,A., Cooke,P., Corum,B., DeArelano,K.,

Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,

Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,

Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,

Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kelle,C., Lander,E., Levine,R.,

Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,

AC115771 204214 bp DNA linear HTG 18-JUN-2003  
Mus musculus clone RP23-116D4, WORKING DRAFT SEQUENCE, 11 ordered  
pieces.

AC115771 GI:31880182

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 204214)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus, clone RP23-116D4

Unpublished

2 (bases 1 to 204214)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barina,N., Baslet,V., Bloom,T., Boguski,M.,

Bouckgeater,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

Chazaro,B., Choepel,Y., Collings,S., Collins,S., Collamore,A.,

Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,

Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kelle,C., Lacroque,K., Lamazares,R.,

Lander,E., Leach,K., Levine,R., Lindblad-Toh,K., Liu,G.,

MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,

McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,

Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,

Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,

Raymond,C., Retta,R., Rieback,M., Riley,K., Rise,C., Rogov,P.,

Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,

Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 204214)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,

Anderson,M., Arachchi,H.M., Barina,N., Bastien,V., Bloom,T.,

Boguski,M., Bouckgeater,B., Camarata,J., Chang,J., Choepel,Y.,

Collamore,A., Cooke,P., Corum,B., DeArelano,K.,

Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,

Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,

Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,

Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kelle,C., Lander,E., Levine,R.,

Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,

AC115771 204214 bp DNA linear HTG 18-JUN-2003  
Mus musculus clone RP23-116D4, WORKING DRAFT SEQUENCE, 11 ordered  
pieces.

AC115771 GI:31880182

HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 204214)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus, clone RP23-116D4

Unpublished

```

misc_feature 136063..181391
             /note="assembly_fragment"
misc_feature 181492..204214
             /note="assembly_fragment"
             clone_end:T7
             vector_side:right"

ORIGIN
Query Match      9.9%;   Score 167.4;   DB 2;   Length 204214;
Best Local Similarity 78.2%;   Pred. NO. 3.2e-38;
Matches 201;   Conservative 0;   Mismatches 56;   Indels 0;   Gaps 0;

QY  838  GAGCTTGCCCTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCGCGGGGTTTGAC 897
Db  142187  GACACCCCTAACCCAGTGGAAAGCCCTCTCGAGCCCTGTGATGAGAGCTGCACAGGGTTGAG 142246

QY  898  TCTCTGTGGAAGACTAGATAGCAGAGAGAGATCAAGAGATCTATGAGAGCTTTAAG 957
Db  142247  TCTGTGCGGGAAGACGGAGTAGTGTGAACACGACGAGTCAAGCGTATCTACGGGGGCTTTAAG 142306

QY  958  AGACAGCGGGGCGGCAAGACACCCATCGCAGGGGTCCCTCCAGTCTCGTGCCTCTGACCATC 1017
Db  142307  AGACAGCAGGCGAAGCACCGCTGGCAGGTGTCTCTCGACACCTCACTGCGGTGACACC 142366

QY  1018  TCATGCCCCAGGGCCACTTCTGTGTGGGGGCGCTGATCCACCCCTGCTGGTGCTCACT 1077
Db  142367  TCCATGCCCCAAGGCCACTTCTGTGTGGGGGCGCCCTGATCCACCCCTGCTGGTGCTCACT 142426

QY  1078  GCTGCCCACTGCAACCGA 1094
Db  142427  GCAGCCACTGTACCGA 142443

```

```

RESULT 15
AC106236/c
LOCUS      Rattus norvegicus clone CH230-131J10, WORKING DRAFT SEQUENCE, 2
DEFINITION      Rattus norvegicus clone CH230-131J10, WORKING DRAFT SEQUENCE, 2
unordered pieces.
ACCESSION      AC106236
VERSION        AC106236.5 GI:30579797
KEYWORDS       HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE         Rattus norvegicus (Norway rat)
ORGANISM       Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE      1 (bases 1 to 224208)
              Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
              Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
              Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
              Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
              Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
              Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
              Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
              Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
              Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
              Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
              Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
              Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
              Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
              Fernandez,S., Finley,M., Flagg,N., Forbes,I., Foster,M., Foster,P.,
              Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,M.,
              Gregorogis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
              Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
              Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
              Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
              Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
              Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
              Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
              Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,D., Li,Z., Liu,J.,
              Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
              Lorenshewa,L., Loulsee,H., Lozado,R.J., Lu,X., Ma,J.,
              Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
              Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
              Mashiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
              Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
              Morgan,M., Morris,K., Morris,S., Muridasa,M., Murphy,M., Nair,L.,
              Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
              Nwankweme,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
              Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
              Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
              Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
              Reilly,B., Reilly,M., Ren,Y., Reuter,M., Rose,R., Ruiz,S.J.,
              Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
              Sanders,W., Savery,G., Scherer,S., Scott,C., Shatsman,S., Shen,H.,
              Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
              Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
              Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
              Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
              Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
              Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
              Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
              Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,D., Yoon,V.,
              Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
              Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
              Weinstock,G. and Gibbs,R.A.
              Direct Submission
Unpublished
2 (bases 1 to 224208)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 224208)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:24942489.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GKCC
Center clone name: CH230-131J10
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 213390 bases at least Q40
Consensus quality: 215800 bases at least Q30
Consensus quality: 216989 bases at least Q20
Estimated insert size: 222031; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

```
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 196130: contig of 196130 bp in length
* 196131 196230: gap of unknown length
* 196231 224208: contig of 27978 bp in length.
FEATURES
    Source
    1..224208
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-131J10"
    2668..3129
    /note="clone boundary"
    clone_end:Sp6
    site:EcoRI
    end_sequence:BH346758"
    200382..201926
    misc_feature
    /note="wgs_contig"
    misc_feature
    200382..201926
    /note="wgs_contig"
    ORIGIN
    Query Match          9.6%;  Score 161;  DB 2;  Length 224208;
    Best Local Similarity 76.7%;  Pred No. 2.8e-36;
    Matches 197;  Conservative 0;  Mismatches 60;  Indels 0;  Gaps 0;
    QY 838 GACGTGCTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCCGGGTTTCAC 897
    Db 71048 GATGCTGCTAACCCATTGGCGACCTTCAGGAACCTGTGATGGAGCTGCCAGGATTCGAC 70989
    QY 898 TCCTGTGGAAGACTGAGATGACAGAGAGAGATCAAGAGAAATCTATGGAGGCTTTAAG 957
    Db 70989 TCCTGCGGGAAGACAGAGATGACTGAAACGCGGTCAAGCGCATCTACGGGGCTTTAAG 70929
    QY 958 AGCAGCGCGGGAAGCAACCCATGGCAGGCGTCCCTCCAGTCCCTGCTGCTGTGACCATC 1017
    Db 70928 AGCAGAGCGGGCAAGCACCCGTGGCAGGTGTCCTCGACACCTCATTGCCGCTGACCACC 70869
    QY 1018 TCCATGCCCCAGGGCCACTTCTGTGTGGGGGCTGATCCACCCCTGCTGGTGTCTCACT 1077
    Db 70868 TCCATGCCCCAGGGCCACTTCTGTGTGGGGGTTCCCTGATTCAACCCCTGCTGGTGTCTCACT 70809
    QY 1078 GCTGCCCACTGCACCGA 1094
    Db 70808 GCAGCCCACTGTACCGA 70792
```

Search completed: May 25, 2004, 09:00:34  
Job time : 4483.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 04:19:43 ; Search time 471 Seconds  
(without alignments)

15179.867 Million cell updates/sec

Title: US-09-912-559-2

Perfect score: 1683

Sequence: 1 atgttgcaggtgtctga.....aaagtgaagtgcttctaa 1683

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_29Jan04.\*

1: Geneseqn1980s.\*

2: Geneseqn1990s.\*

3: Geneseqn2000s.\*

4: Geneseqn2001as.\*

5: Geneseqn2001bs.\*

6: Geneseqn2002s.\*

7: Geneseqn2003as.\*

8: Geneseqn2003bs.\*

9: Geneseqn2003cs.\*

10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1683	100.0	1683	6	AAL45697 Human blo
2	1679.8	99.8	1683	6	AAL45696 Human blo
3	1679.8	99.8	3008	6	ABN95170 Gene #166
4	1679.8	99.8	3008	7	ACC51204 Human Plk
5	1675	99.5	3623	3	AAc76693 Human ORF
6	526.4	31.3	617	3	AAa44763 Human sec
7	368.4	21.9	428	8	ACH21030 Human adu
8	241	14.3	397	6	ABi9281 Mouse isc
9	120.2	7.1	451	4	ABa58323 Human fce
10	120.2	7.1	451	4	AAi37947 Probe #66
11	120.2	7.1	451	4	AAk32096 Human bon
12	120.2	7.1	451	4	AAk06420 Human bra
13	120.2	7.1	451	4	ABs31791 Human liv
14	120.2	7.1	451	6	ABs06863 Human gen
15	117	7.0	117	4	ABa70919 Human fce
16	117	7.0	117	4	AAi51106 Probe #19
17	117	7.0	117	4	AAk45153 Human bon
18	117	7.0	117	4	AAk19188 Human liv
19	117	7.0	117	4	ABs44824 Human liv
20	117	7.0	117	6	ABs19403 Human gen
21	114.4	6.8	2033	2	AAq63951 Hepatocyt
22	114.4	6.8	2036	6	ABi65242 Lung canc
23	114.4	6.8	2036	6	ABn95114 Gene #161

## ALIGNMENTS

RESULT 1

AAL45697  
ID AAL45697 standard; DNA; 1683 BP.

XX  
AC AAL45697;

XX  
DT 13-JUN-2002 (first entry)

XX  
DE Human blood coagulation factor VII activating protease mutant DNA.

XX  
KW Human; blood coagulation factor VII activating protease; FSAP;  
single-chain plasminogen activator; bleeding disorder; haematological;  
haemostatic; mutant; gene; ds.

XX  
OS Homo sapiens.  
Synthetic.

XX  
FH Key Location/Qualifiers  
CDS 1..1683

FT /\*tag= a  
FT /product= "mutant FSAP"

XX  
PN EP1182258-A1.

XX  
PD 27-FEB-2002.

XX  
PF 05-JUL-2001; 2001EP-00115691.

XX  
PR 26-JUL-2000; 2000DE-01036641.

XX  
PR 10-OCT-2000; 2000DE-01050040.

XX  
PR 21-OCT-2000; 2000DE-01052319.

XX  
PR 12-APR-2001; 2001DE-01018706.

XX  
PA (AVET ) AVENTIS BEHRING GMBH.

XX  
PI Roemisch J, Stoehr H, Feussner A, Lang W, Weimer T, Becker M;  
Nerlich C, Muth-Naumann G;

XX  
PI WPI; 2002-270939/32.

XX  
DR P-PSDB; AA017145.

XX  
PT New nucleic acid encoding mutant factor 7 activating protease, useful for  
diagnosis, treatment and prevention of coagulation disorders, also  
related protein and antibodies.

XX  
PS Disclosure; Page 16-17; 27pp; German.

Ade06939 Novel cod  
Abx17688 DNA encod  
Aaq63945 Gene comp  
Aaq05673 v-PA alph  
Aaq12867 JM1-229 c  
Adb52711 Primary r  
Aan70992 Modified  
Aan70991 Modified  
Aaq05675 v-PA\_beta  
Aaq05674 v-PA alph  
Aaq00543 Vampire b  
Aan90048 Chicken u  
Aan91217 Mutated c  
Aan12071 T-PA wic  
Aaq12074 T-PA vari  
Aad27741 Human ful  
Aaq79004 Gene comp  
Aaq40658 tPA-2 CDN  
Aat87456 mt-PA6-E.  
Aat62606 DNA encod  
Aaq53322 Human tru  
Aaq53320 Human tru

24 99.4 5.9 2145 9 ADE06939  
25 93.2 5.5 2299 7 ABX17688  
26 93 5.5 970 2 Aaq63945  
27 87.6 5.2 2252 2 Aaq05673  
28 87.4 5.2 2512 2 Aaq12867  
29 87.2 5.2 2445 9 Adb52711  
30 86.8 5.2 1323 1 Aan70992  
31 85.2 5.1 1422 1 Aan70991  
32 84.8 5.0 2137 2 Aaq05675  
33 79.6 4.7 2257 2 Aaq05674  
34 78 4.6 1620 2 Aaq00543  
35 77.2 4.6 919 2 Aat90048  
36 75.4 4.5 1689 1 Aan91217  
37 73.8 4.4 2097 2 Aaq12071  
38 73.8 4.4 2100 2 Aaq12074  
39 72.8 4.3 614 6 Aad27741  
40 72.2 4.3 329 2 Aaq79004  
41 72.2 4.3 1047 2 Aaq40658  
42 72.2 4.3 1065 2 Aat87456  
43 72.2 4.3 1065 2 Aat62606  
44 72.2 4.3 1068 2 Aaq53322  
45 72.2 4.3 1068 2 Aaq53320

QY 1441 CAACCTCTATGACACATGATTGATGACAGATATGATCTGTGACGAAATCTTCAGAAACCT 1500  
 Db 1441 CAACTCTATGACACATGATTGATGACAGATATGATCTGTGACGAAATCTTCAGAAACCT 1500  
 QY 1501 GGGCAAGACACCTGCGAGGTGACTCTGAGAGGCCCTGACCTGTGAGAGGACGCGACC 1560  
 Db 1501 GGGCAAGACACCTGCCAGGTGACTCTGAGAGGCCCTGACCTGTGAGAGGACGCGACC 1560  
 QY 1561 TACTACCTCTATGGGATAGTGTGAGCTGGGGCTGTGAGTGTGAGAGAGGCCAGGGGCTTAC 1620  
 Db 1561 TACTACCTCTATGGGATAGTGTGAGCTGGGGCTGTGAGTGTGAGAGAGGCCAGGGGCTTAC 1620  
 QY 1621 ACCCAAGTTACCAATTCCTGAATTTGGATCAAGGCCACCAATCAAAAGTGAAGTGGCTTC 1680  
 Db 1621 ACCCAAGTTACCAATTCCTGAATTTGGATCAAGGCCACCAATCAAAAGTGAAGTGGCTTC 1680  
 QY 1681 TAA 1683  
 Db 1681 TAA 1683

RESULT 3

ABN95170  
 ID ABN95170 standard; DNA; 3008 BP.

XX AC ABN95170;

XX 13-AUG-2002 (first entry)

XX Gene #1668 used to diagnose liver cancer.

XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
 XX metastatic liver tumor; cytostatic; expression profiler; disease state;  
 XX disease progression; drug toxicity; drug efficacy; drug metabolism.

XX Homo sapiens.

XX WO200229103-A2.

XX 11-APR-2002.

XX 02-OCT-2001; 2001WO-US030589.

XX 02-OCT-2000; 2000US-0237054P.

XX (GENE-) GENE LOGIC INC.

XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX WPI; 2002-426119/45.

XX Diagnosing and detecting the progression of liver cancer, hepatocellular  
 XX carcinoma or metastatic liver tumor in a patient, involves detecting the  
 XX level of expression of two or more genes in a liver tissue sample.

XX Claim 1; SEQ ID NO 1668; 298pp; English.

XX The invention relates to a novel method for diagnosing and detecting the  
 XX progression of liver cancer, hepatocellular carcinoma or metastatic liver  
 XX tumour in a patient, and differentiating metastatic liver cancer from  
 XX hepatocellular carcinoma in a patient, involving detecting the level of  
 XX expression of two or more genes represented in ABN93503-ABN97455 in a  
 XX tissue sample. The method of the invention has hepatotropic, and  
 XX cytostatic activity. The method is useful for diagnosing and detecting  
 XX the progression of liver cancer, hepatocellular carcinoma and metastatic  
 XX liver carcinoma in a patient. The method is useful for identifying  
 XX expression profiles which serve as useful diagnostic markers as well as  
 XX markers that can be used to monitor disease states, disease progression,  
 XX drug toxicity, drug efficacy and drug metabolism. Note: The sequence data  
 XX for this patent did not form part of the printed specification, but was  
 XX obtained in electronic format directly from WIPO at  
 XX ftp.wipo.int/pub/published\_pat\_sequences

SQ Sequence 3008 BP; 791 A; 837 C; 703 G; 677 T; 0 U; 0 Other;  
 Query Match 99.8%; Score 1679.8; DB 6; Length 3008;  
 Best local Similarity 99.9%; Pred. No. 0;  
 Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 ATGTTTGCCAGGATGTCTGATCTCCATGTTCTGTGTTAATGGCTCTGGTGGGAAAGACA 60  
 Db 97 ATGTTTGCCAGGATGTCTGATCTCCATGTTCTGTGTTAATGGCTCTGGTGGGAAAGACA 156  
 QY 61 GCCTGTGGGTTCTCCCTGATGTCTTTATTTGGAAGCCTGGACCCAGATGGACCCCTGAC 120  
 Db 157 GCCTGTGGGTTCTCCCTGATGTCTTTATTTGGAAGCCTGGACCCAGATGGACCCCTGAC 216  
 QY 121 CAGTATGATTACAGTACGAGGATTAATAATCAGGAAGAGAACACACAGTAGCACACTTACC 180  
 Db 217 CAGTATGATTACAGTACGAGGATTAATAATCAGGAAGAGAACACACAGTAGCACACTTACC 276  
 QY 181 CATGCTGAGAATCTGACTGTGATCTACTGAGGACCAAGCTGATCTCATGCGCCCAAC 240  
 Db 277 CATGCTGAGAATCTGACTGTGATCTACTGAGGACCAAGCTGATCTCATGCGCCCAAC 336  
 QY 241 CCCTGTGAACACGGTGGGACTGCTCGTCCATGGGAGCACCTTTCACATGAGCTGCGTG 300  
 Db 337 CCCTGTGAACACGGTGGGACTGCTCGTCCATGGGAGCACCTTTCACATGAGCTGCGTG 396  
 QY 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGAATAATAGTGCAGAGGACCAACCCATGT 360  
 Db 397 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGAATAATAGTGCAGAGGACCAACCCATGT 456  
 QY 361 GGCGGGGCCAATGTCTCATTACCCAGAGTCTCTCTACTACCCGCTGTGTCTGTAACAC 420  
 Db 457 GGCGGGGCCAATGTCTCATTACCCAGAGTCTCTCTACTACCCGCTGTGTCTGTAACAC 516  
 QY 421 CTTTACACAGTCCGAGCTGCTCCCAAGTGTCTCTGTATGACAGGACCAACCCCTGCGAG 480  
 Db 517 CTTTACACAGTCCGAGCTGCTCCCAAGTGTCTCTGTATGACAGGACCAACCCCTGCGAG 576  
 QY 481 AATGGGGTACTGCTCCCGGCATAAGCGGAGATCCAAAGTTCACTGTGCTGTGCTCCGAC 540  
 Db 577 AATGGGGTACTGCTCCCGGCATAAGCGGAGATCCAAAGTTCACTGTGCTGTGCTCCGAC 636  
 QY 541 CAGTTCAAGGGGAAATCTGTGAAATAGTTTCTCATGCTGTATGTTGGGATGGCTAC 600  
 Db 637 CAGTTCAAGGGGAAATCTGTGAAATAGTTTCTCATGCTGTATGTTGGGATGGCTAC 696  
 QY 601 TCTTACCGAGGAAATGATAGGACAGTCAACAGCATGCGTCCCTTACTGGAATCC 660  
 Db 697 TCTTACCGAGGAAATGATAGGACAGTCAACAGCATGCGTCCCTTACTGGAATCC 756  
 QY 661 CACTCTCTTTGCGAGGAAATTAACAATGTTTATGAGGATGCTGAAACCCATGGGATT 720  
 Db 757 CACTCTCTTTGCGAGGAAATTAACAATGTTTATGAGGATGCTGAAACCCATGGGATT 816  
 QY 721 GGGGAACACAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTGCTGTTTATTA 780  
 Db 817 GGGGAACACAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTGCTGTTTATTA 876  
 QY 781 GTTACCAATGACAGGTGAAATGGAATACTGTGATGTCAGCTGCTAGCCAGGAC 840  
 Db 877 GTTACCAATGACAGGTGAAATGGAATACTGTGATGTCAGCTGCTAGCCAGGAC 936  
 QY 841 GTTGCTTACCGAGGAAAGCCCACTGAGCCATCAACAGCTTCCGGGTTTGAATCC 900  
 Db 937 GTTGCTTACCGAGGAAAGCCCACTGAGCCATCAACAGCTTCCGGGTTTGAATCC 996  
 QY 901 TGTGAAAGACTGAGATAGCAGAGGAGGAGATCAAGAGAAATCTATGGAGGCTTTAAGAGC 960  
 Db 997 TGTGAAAGACTGAGATAGCAGAGGAGGAGATCAAGAGAAATCTATGGAGGCTTTAAGAGC 1056  
 QY 961 ACGGGGGCAGACCCCATGAGGCGCTCCCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 Db 1057 ACGGGGGCAGACCCCATGAGGCGCTCCCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1116

QY 1021 ATGCCCGAGGCGCACTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080  
Db ATGCCCGAGGCGCACTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1176  
QY 1081 GCCCACTGCACCGCATATAAAACACAGACATCTAAAGTGTGTAGGGGACCAAGGACCTG 1140  
Db GCCCACTGCACCGCATATAAAACACAGACATCTAAAGTGTGTAGGGGACCAAGGACCTG 1236  
QY 1141 AAGAAAGAGAAATTTCTATGACAGAGCTTTAGGGTGAGAGATATTTCAAGTACAGCCAC 1200  
Db AAGAAAGAGAAATTTCTATGACAGAGCTTTAGGGTGAGAGATATTTCAAGTACAGCCAC 1296  
QY 1201 TACAATGAAGAGATGAGATTTCCCAACATGATATTCATTTGCTCAAGTTTAAAGCCAGTG 1260  
Db TACAATGAAGAGATGAGATTTCCCAACATGATATTCATTTGCTCAAGTTTAAAGCCAGTG 1356  
QY 1261 GATGGTCACTGTGCTTAGAATCCAAATACGTGAAGACTGTGTGCTTGGCTGTGATGGGTCC 1320  
Db GATGGTCACTGTGCTTAGAATCCAAATACGTGAAGACTGTGTGCTTGGCTGTGATGGGTCC 1416  
QY 1321 TTTCCCTCTGGAGTGAGTGCCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1380  
Db TTTCCCTCTGGAGTGAGTGCCACATCTCTGGCTGGGGTGTACAGAAACAGGAAAGGG 1476  
QY 1381 TCCGCGCAGCTCTGGATGCCAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC 1440  
Db TCCGCGCAGCTCTGGATGCCAAAGTCAAGCTGATTCGCAACACTTTGTGCAACTCCCGC 1536  
QY 1441 CAATCTATGACACATGATTTGATGACAGTATGATCTGTCAGGAATCTTCAGAACCT 1500  
Db CAATCTATGACACATGATTTGATGACAGTATGATCTGTCAGGAATCTTCAGAACCT 1596  
QY 1501 GGGCAAGACACTGCCAGGGTGAATCTCTGGAGGCCCTGACCTGTGAGAGGACGGCACC 1560  
Db GGGCAAGACACTGCCAGGGTGAATCTCTGGAGGCCCTGACCTGTGAGAGGACGGCACC 1656  
QY 1561 TACTACGTCTATGGATAGTGAAGTGGGGCTGAGTGTGAGAGAGCCAGGGGTCTAC 1620  
Db TACTACGTCTATGGATAGTGAAGTGGGGCTGAGTGTGAGAGAGCCAGGGGTCTAC 1716  
QY 1621 ACCCAAGTTACCAATTCCTGAATTTGGATCAAGCCACCATCAAAAGTGAAGTGGGCTTC 1680  
Db ACCCAAGTTACCAATTCCTGAATTTGGATCAAGCCACCATCAAAAGTGAAGTGGGCTTC 1776  
QY 1681 TAA 1683  
Db 1777 TAA 1779  
RESULT 4  
ACCS1204  
ID ACCS1204 standard; cDNA; 3008 BP.  
AC ACCS1204;  
XX  
XX  
DE 16-JUN-2003 (first entry)  
DE Human Plk-1 related cDNA sequence hmft-0306 SEQ ID NO:89.  
XX  
XX Human; hepatoblastoma; cancer detection probe; cancer; detection;  
KW hepatocellular carcinoma; hereditary non-polyposis colorectal cancer;  
KW desmoid tumour; anaplastic thyroid carcinoma; Wilm's tumour; tumour;  
XX Plk-1; polo-like kinase-1; gene; ss.  
OS Homo sapiens.  
XX  
XX WO2003018807-A1.  
XX  
XX 06-MAR-2003.  
XX  
XX 26-AUG-2002; 2002WO-JP008580.  
XX

PR 24-AUG-2001; 2001JP-00255225.  
XX (HISM) HISAMITSU PHARM CO LTD.  
PA (CHIB-) CHIBA PREFECTURE.  
XX  
XX Nakagawara A;  
PI  
XX WPI; 2003-268424/26.  
XX  
XX Nucleic acid sequences differently expressed between hepatoblastoma and  
PT normal liver tissue, are useful for cancer detection and diagnosis.  
XX  
XX Claim 4; Page 142-144; 180pp; Japanese.  
XX  
XX The present invention describes nucleic acid sequences (I) having a  
CC different degree of expression in hepatoblastoma from their expression in  
CC normal liver tissue. ACCS1116 to ACCS1219 represents specifically claimed  
CC examples of (I). Also described: (1) nucleic acids stringently  
CC hybridizing to (I); (2) cancer detection probes containing one or more of  
CC 104 listed sequences (II), or partial sequences derived from them; (3)  
CC see ACCS1116 to ACCS1194), or partial sequences derived from them; (3)  
CC PCR primers for cancer detection based on sequences (II); (4) marker  
CC proteins for cancer detection, encoded by (II); (5) diagnostic reagents  
CC for cancer diagnosis, containing (II) or their partial sequences. The  
CC nucleic acid sequences are useful in the detection and diagnosis of  
CC cancers including liver, colon, breast, kidney, bladder, ovary and  
CC thyroid cancer, especially for hepatoblastoma, hepatocellular carcinoma,  
CC hereditary non-polyposis colorectal cancer, desmoid tumour, anaplastic  
CC thyroid carcinoma and Wilm's tumour. They are also used as markers for  
CC predicting the prognosis of these tumours. ACCS1220 to ACCS1233 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC nucleic acid sequences given in ACCS1116 to ACCS1219 are related to human  
CC Plk-1 (polo-like kinase-1), which is located on chromosome 16p12  
XX  
XX Sequence 3008 BP; 791 A; 837 C; 703 G; 677 T; 0 U; 0 Other;  
Query Match 99.8%; Score 1679.8; DB 7; Length 3008;  
Best Local Similarity 99.9%; Pred No. 0;  
Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ATGTTTGGCAGATGTCGATCTCCATGTTCTGTGTTTAAATGGCTCTGGTGGAAAGACA 60  
Db 97 ATGTTTGGCAGATGTCGATCTCCATGTTCTGTGTTTAAATGGCTCTGGTGGAAAGACA 156  
QY 61 GCCTGTGGGTTCTCCCTGATGTTCTTATTGGAAAGCTGGACCAGACTGGACCCCTGAC 120  
Db 157 GCCTGTGGGTTCTCCCTGATGTTCTTATTGGAAAGCTGGACCAGACTGGACCCCTGAC 216  
QY 121 CAGTATGATTACAGCTACGAGGATTATATATCAGGAAGAGAACACCACTAGCACACTTACC 180  
Db 217 CAGTATGATTACAGCTACGAGGATTATATATCAGGAAGAGAACACCACTAGCACACTTACC 276  
QY 181 CATGCTGAGAACTCTGACTGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 240  
Db 277 CATGCTGAGAACTCTGACTGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC 336  
QY 241 CCCTGTGAACACGGTGGGACTGCTCTGCTCCATGGAGCACCCTTACATGACAGTGCCTG 300  
Db 337 CCCTGTGAACACGGTGGGACTGCTCTGCTCCATGGAGCACCCTTACATGACAGTGCCTG 396  
QY 301 GTCCTTTCTCTGGGAAATAGTGTGAGAAAGTGAATAATACGTGCAAGGACCAACCATCT 360  
Db 397 GTCCTTTCTCTGGGAAATAGTGTGAGAAAGTGAATAATACGTGCAAGGACCAACCATCT 456  
QY 361 GGCCGGGGCCCAATGCTCTCATTTACCCAGAGTCTCTCCCTACTACCGTGTGTGTAAACAC 420  
Db 457 GGCCGGGGCCCAATGCTCTCATTTACCCAGAGTCTCTCCCTACTACCGTGTGTGTAAACAC 516  
QY 421 CCTTACACAGGTCCTGCTGCTCCCAAGTGGTTCTGTATGAGGCGCAACCCCTGCCAG 480  
Db 517 CCTTACACAGGTCCTGCTGCTCCCAAGTGGTTCTGTATGAGGCGCAACCCCTGCCAG 576  
QY 481 AATGGGGCTACCTGCTCCCGGATTAAGCGGAGATCCAAAGTTACCTGTGCTGTCCCGAC 540



Db 577 AATGGGCTACCTGCTCCCGGCAATAAGCGAGATCCAAAGTTTCACTGTGCTGTCCCGAC 636  
Qy 541 CAGTTCAAGGGGAAATCTGTGAATAGGTTCTGATGCTGTATGCTGTGCGCATGGCTAC 600  
Db 637 CAGTTCAAGGGGAAATCTGTGAATAGGTTCTGATGCTGTATGCTGTGCGCATGGCTAC 696  
Qy 601 TCTTACCGAGGGAATGAATAGGACAGTCAACAGCATGCGTGCTGCTTTACTGGAATCC 660  
Db 697 TCTTACCGAGGGAATGAATAGGACAGTCAACAGCATGCGTGCTGCTTTACTGGAATCC 756  
Qy 661 CAGCTCTCTTGAGAGGAATTAACAATGTTTATGGAGATGCTGAAACCCATGGGATT 720  
Db 757 CAGCTCTCTTGAGAGGAATTAACAATGTTTATGGAGATGCTGAAACCCATGGGATT 816  
Qy 721 GGGGAACACAAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTGCTTTATATAA 780  
Db 817 GGGGAACACAAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTGCTTTATATAA 876  
Qy 781 GTTACCAATGACAGGTGAATGGGAATGCTGTGATGCTCAGCCTGCTCAGCCAGGAC 840  
Db 877 GTTACCAATGACAGGTGAATGGGAATGCTGTGATGCTCAGCCTGCTCAGCCAGGAC 936  
Qy 841 GTTGCTTACCAGAGGAAGCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC 900  
Db 937 GTTGCTTACCAGAGGAAGCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC 996  
Qy 901 TGTGGAAGACTGAGATAGCAGAGAGGAAGATCAAGAGATCTATGAGGCTTTAGAGC 960  
Db 997 TGTGGAAGACTGAGATAGCAGAGAGGAAGATCAAGAGATCTATGAGGCTTTAGAGC 1056  
Qy 961 ACGCGGGCAAGCAACCATGCGAGGCTCCTCCAGTCTGCTGCTGCTGACCATCTCC 1020  
Db 1057 ACGCGGGCAAGCAACCATGCGAGGCTCCTCCAGTCTGCTGCTGCTGACCATCTCC 1116  
Qy 1021 ATGCCCCAGGGCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080  
Db 1117 ATGCCCCAGGGCACTTCTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1176  
Qy 1081 GCCCACTGCACCGACATAAAACAGAGATCTAAAGTGTGTGTAGGGACACAGGACCTG 1140  
Db 1177 GCCCACTGCACCGACATAAAACAGAGATCTAAAGTGTGTGTAGGGACACAGGACCTG 1236  
Qy 1141 AAGAAAGAAATTTATGAGCAGAGGTTTAGGTTGAGAGATATTCAGATGACGCCAC 1200  
Db 1237 AAGAAAGAAATTTATGAGCAGAGGTTTAGGTTGAGAGATATTCAGATGACGCCAC 1296  
Qy 1201 TACATGAAGAGATGATTCCTCCCAATGATATTCATGCTCAAGTTTAAAGCCAGTG 1260  
Db 1297 TACATGAAGAGATGATTCCTCCCAATGATATTCATGCTCAAGTTTAAAGCCAGTG 1356  
Qy 1261 GATGTCACCTGCTCTAGAAATCAAAATACGTGAAGACTGTGTGCTTGCCTGATGGGTCC 1320  
Db 1357 GATGTCACCTGCTCTAGAAATCAAAATACGTGAAGACTGTGTGCTTGCCTGATGGGTCC 1416  
Qy 1321 TTTCCCTCTGGAGTGGCCATCTCTGGCTGGGGTGTACAGAAACAGAAAGGG 1380  
Db 1417 TTTCCCTCTGGAGTGGCCATCTCTGGCTGGGGTGTACAGAAACAGAAAGGG 1476  
Qy 1381 TCCCGCCAGCTCCTGGATGCCAAAGTCAAGCTGATGTCACACACTTTGTGCAACTCCCGC 1440  
Db 1477 TCCCGCCAGCTCCTGGATGCCAAAGTCAAGCTGATGTCACACACTTTGTGCAACTCCCGC 1536  
Qy 1441 CAACCTATGACCAATGATGATGACAGATGATCTGTGCGAATCTTCAGAAACCT 1500  
Db 1537 CAACCTATGACCAATGATGATGACAGATGATCTGTGCGAATCTTCAGAAACCT 1596  
Qy 1501 GGGCAAGACACTGCGAGGGTGACTCTGGAGGCCCTGACCTGTGAGAGGAGCGGACCC 1560  
Db 1597 GGGCAAGACACTGCGAGGGTGACTCTGGAGGCCCTGACCTGTGAGAGGAGCGGACCC 1656  
Qy 1561 TACTACGCTTATGGGATGATGATGCTGGGGCTGAGATGTGAGAGAGCGGCCGGGTCTAC 1620

Db 1657 TACTACGCTTATGGATAGTGAAGTGGGGCTGGAGTGTGGGAAGAGCGCGGTCTAC 1716  
Qy 1621 ACCCAAGTTACCAAAATTCCTGAATTTGGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC 1680  
Db 1717 ACCCAAGTTACCAAAATTCCTGAATTTGGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC 1776  
Qy 1681 TAA 1683  
Db 1777 TAA 1779  
RESULT 5  
AAC76693  
ID AAC76693 standard; cDNA; 3623 BP.  
XX  
AC AAC76693;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
Human ORFX ORF2248 polynucleotide sequence SEQ ID NO:4495.  
DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
XX  
KW Human; antipsoriatic; antiparkinsonian; nontropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; erythematoidem; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200058473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000; 2000WO-US008621.  
XX  
PR 31-MAR-1999; 98US-0127607P.  
PR 02-APR-1999; 98US-0127636P.  
PR 05-APR-1999; 98US-0127728P.  
PR 30-MAR-2000; 2000US-00540763.  
XX  
PA (CURA-) CUPAGEN CORP.  
XX  
PI Shinketsu RA, Leach M;  
XX  
DR WPI; 2000-602362/57.  
XX  
DR P-PSDB; AAB42484.  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease.  
XX  
PS Claim 5; Page 3679-3681; 5507pp; English.  
XX  
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;  
CC antipsoriatic; antiparkinsonian; nontropic; neuroprotective; osteopathic;  
CC anticonvulsant; antirheumatic; immunosuppressant; immunostimulant;  
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;  
CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The  
CC sequences can be used for determining the presence of or predisposition  
CC to, or preventing or treating pathological conditions associated with an  
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX.

CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
CC used to treat cancers, proliferative disorders, neurodegenerative  
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
CC (SCID), AIDS, viral bacterial or fungal infection, malaria, autoimmune  
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
CC cartilage damage, nocturnal haemoglobinuria, anti-inflammatory disease; to  
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive  
XX

SQ Sequence 3623 BP; 848 A; 1005 C; 930 G; 840 T; 0 U; 0 Other;

Query Match 99.5%; Score 1675; DB 3; Length 3623;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1678; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGTTTGCAGGATCTCTGATCTCCATGTTCTGCTTTAATGGCTTGTGGTGGAAAGACA 60  
DB 97 ATGTTTGCAGGATCTCTGATCTCCATGTTCTGCTTTAATGGCTTGTGGTGGAAAGACA 156  
QY 61 GCCTGTGGTCTCCCTGATGCTTTATTGAAAGCCTGGACCCAGACTGACCCCTGAC 120  
DB 157 GCCTGTGGTCTCCCTGATGCTTTATTGAAAGCCTGGACCCAGACTGACCCCTGAC 216  
QY 121 CAGTATGATTACAGCTACGAGGATTTAATCAGGAAGAGAACACAGTAGCACACTTACC 180  
DB 217 CAGTATGATTACAGCTACGAGGATTTAATCAGGAAGAGAACACAGTAGCACACTTACC 276  
QY 181 CATGCTGAGATCTCTGATCTGATCTACACTGAGGACCAAGCTGATCCATGCGAGCCCAAC 240  
DB 277 CACGCTGAGATCTCTGATCTGATCTACACTGAGGACCAAGCTGATCCATGCGAGCCCAAC 336  
QY 241 CCCTGTGAACACGGTGGGAGCTGCTCTCATGGGAGCACCTTACATGAGCTGCTG 300  
DB 337 CCCTGTGAACACGGTGGGAGCTGCTCTCATGGGAGCACCTTACATGAGCTGCTG 396  
QY 301 GCTCTTCTCTGGAAATAGTGTGAGAAAGTGCAAAATAGCTGCAAGGACAAACCCATGT 360  
DB 397 GCTCTTCTCTGGAAATAGTGTGAGAAAGTGCAAAATAGCTGCAAGGACAAACCCATGT 456  
QY 361 GGCCTGGGCGCAATGCTCATTAACCCAGAGTCTCTCTACTACCGCTGCTGTAAACAC 420  
DB 457 GGCCTGGGCGCAATGCTCATTAACCCAGAGTCTCTCTACTACCGCTGCTGTAAACAC 516  
QY 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGTCTCTGATGAGGCCAAAACCCCTGCGAG 480  
DB 517 CTTTACAGAGTCCAGCTGCTCCCAAGTGTCTCTGATGAGGCCAAAACCCCTGCGAG 576  
QY 481 AATGGGGTCACTGCTCCCGCATAGCGGAGATCCAAAGTTCACTGCTGCTGCCGAC 540  
DB 577 AATGGGGTCACTGCTCCCGCATAGCGGAGATCCAAAGTTCACTGCTGCTGCCGAC 636  
QY 541 CAGTTCAAGGGGAAATCTGTGAATAGGTTCTGATGACTCTATGTTGGCGATGGCTAC 600  
DB 637 CAGTTCAAGGGGAAATCTGTGAATAGGTTCTGATGACTCTATGTTGGCGATGGCTAC 696  
QY 601 TCTTACCGGAGAAATGAAATAGGACAGTCAACAGCATGCTGCTTACTGGAATCC 660  
DB 697 TCTTACCGGAGAAATGAAATAGGACAGTCAACAGCATGCTGCTTACTGGAATCC 756  
QY 661 CACCTCCTCTTCAGGAGAAATACAACTGTTTATGGAGGATGCTGAAACCCATGGGAT 720  
DB 757 CACCTCCTCTTCAGGAGAAATACAACTGTTTATGGAGGATGCTGAAACCCATGGGAT 816  
QY 721 GGGGACACAAATTTCTGAGAAACCCAGATGGGACGAAAGCCCTGTGCTTTATTA 780  
DB 817 GGGGACACAAATTTCTGAGAAACCCAGATGGGACGAAAGCCCTGTGCTTTATTA 876  
QY 781 GTTACCAATGACAGGTGAAATGGGAAATCTGTGATGCTCTAGCCTGCTAGCCCGAGAC 840  
DB 877 GTTACCAATGACAGGTGAAATGGGAAATCTGTGATGCTCTAGCCTGCTAGCCCGAGAC 936  
QY 841 GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCCGGGGTTTGACTCC 900

DB 937 GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCCGGGGTTGACTCC 996  
QY 901 TGTGGAAGACCTGAGATACAGAGAGAGAGATCAAGAGAAATCTATGGAGCTTTAAGAC 960  
DB 997 TGTGGAAGACCTGAGATACAGAGAGAGAGATCAAGAGAAATCTATGGAGCTTTAAGAC 1056  
QY 961 ACGGCGGCAAGCACCATGGAGGCTCCCTCCAGTCTCCCTGCTGCTGACCATCTCC 1020  
DB 1057 ACGGCGGCAAGCACCATGGAGGCTCCCTCCAGTCTCCCTGCTGCTGACCATCTCC 1116  
QY 1021 ATGCCCCAGGGCCACTTCTGTGTGGGGGCTGATCCACCCCTGTGGGTGCTCACTGCT 1080  
DB 1117 ATGCCCCAGGGCCACTTCTGTGTGGGGGCACTGATCCACCCCTGTGGGTGCTCACTGCT 1176  
QY 1081 GCCCACTGCACCGACATAAATAAACACAGACATCTAAAGTGCTGTAGGGGACACAGGACCTG 1140  
DB 1177 GCCCACTGCACCGACATAAATAAACACAGACATCTAAAGTGCTGTAGGGGACACAGGACCTG 1236  
QY 1141 AAGAAAGAGAAATTTTCATGAGCAGAGCTTTTAAAGGTGCAAGATATTCAGGTACAGCCAC 1200  
DB 1237 AAGAAAGAGAAATTTTCATGAGCAGAGCTTTTAAAGGTGCAAGATATTCAGGTACAGCCAC 1296  
QY 1201 TACAATGAAGAGATGAGATTTCCCAATGATATTTGCAATTTGCTCAAGTTAAAGCCAGTG 1260  
DB 1297 TACAATGAAGAGATGAGATTTCCCAATGATATTTGCAATTTGCTCAAGTTAAAGCCAGTG 1356  
QY 1261 GATGTCATCTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTTGCCTGATGCTGCTCC 1320  
DB 1357 GATGTCATCTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTTGCCTGATGCTGCTCC 1416  
QY 1321 TTTCCCTCTGGGAGTGAGTGCCACATCTCTGCTGGGGTGTACAGAAACAGGAAAGGG 1380  
DB 1417 TTTCCCTCTGGGAGTGAGTGCCACATCTCTGCTGGGGTGTACAGAAACAGGAAAGGG 1476  
QY 1381 TCCCGCAGCTCTGATGCTCAAGCTGATTTGCCAATCTTGTGCACTTGTGCACTCCCGC 1440  
DB 1477 TCCCGCAGCTCTGATGCTCAAGCTGATTTGCCAATCTTGTGCACTTGTGCACTCCCGC 1536  
QY 1441 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
DB 1537 CAACTCTATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1596  
QY 1501 GGCACAGACACTGCCAGGGTGACCTGAGAGCCCTGACCTGTGAGAGAGAGAGAGAGAGAGAG 1560  
DB 1597 GGCACAGACACTGCCAGGGTGACCTGAGAGCCCTGACCTGTGAGAGAGAGAGAGAGAGAGAG 1656  
QY 1561 TACTAGCTCTATGGGATAGTGAGCTGGGGCTGGAGTGTGAGAAAGGAGGAGGAGGAGGAGGAG 1620  
DB 1657 TACTAGCTCTATGGGATAGTGAGCTGGGGCTGGAGTGTGAGAAAGGAGGAGGAGGAGGAG 1716  
QY 1621 ACCCAAGTTACAAATTTCTGAAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680  
DB 1717 ACCCAAGTTACAAATTTCTGAAATGGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1776  
QY 1681 TAA 1683  
DB 1777 TAA 1779

RESULT 6

AAA44763

ID AAA44763 standard; cdna; 617 BP.

XX AAA44763;

DT 21-AUG-2000 (first entry)

XX Human secreted expressed sequence tag SEQ ID NO:1338.

XX Human; mouse; chicken; rat; secreted expressed sequence tag; EST;

XX expressed sequence tag; EST; probe; chemotactic; proliferative;

XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;



CC are useful in diagnostics as expressed sequence tags (EST) for  
CC identifying expressed genes or for physical mapping of the human genome,  
CC in forensics, in assessing biodiversity, or in identifying mutations  
CC responsible for genetic disorders and other traits. The nucleotide  
CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
CC is useful for generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030073623  
XX

SQ Sequence 428 BP; 119 A; 108 C; 105 G; 96 T; 0 U; 0 Other;

Query Match 21.9%; Score 368.4; DB 8; Length 428;  
Best Local Similarity 99.7%; Pred. No. 1.6e-94;  
Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGTTTCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGGTGGAAAGACA 60  
DB 59 ATGTTTCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGCTCTGGTGGAAAGACA 118  
QY 61 GCCTGTGGTTCCTCCGTGCTGTTTATGGAAGCCTGGACCCAGACTGGACCCCTGAC 120  
DB 119 GCCTGTGGTTCCTCCGTGCTGTTTATGGAAGCCTGGACCCAGACTGGACCCCTGAC 178  
QY 121 CAGTATGATTACACTACGAGGATTAATAACAGGAAGACACACAGTACCACTTACC 180  
DB 179 CAGTATGATTACACTACGAGGATTAATAACAGGAAGACACACAGTACCACTTACC 238  
QY 181 CATGCTGAGAAATCCTGATCTGTACTACACTGAGACCAAGCTGATCCATCCAGCCCAAC 240  
DB 239 CATGCTGAGAAATCCTGATCTGTACTACACTGAGACCAAGCTGATCCATCCAGCCCAAC 298  
QY 241 CCCTGTGAACACGGTGGGACTGCTCGTCCATGGGAGCACCTTCACATGACGTGCTG 300  
DB 299 CCCTGTGAACACGGTGGGACTGCTCGTCCATGGGAGCACCTTCACATGACGTGCTG 358  
QY 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAATGCAAAATACGTGCAAGGCAACCCATGT 360  
DB 359 GCTCCTTTCTCTGGGAATAGTGTGAGAAATGCAAAATACGTGCAAGGCAACCCATGT 418  
QY 361 GCGCCGGGCC 370  
DB 419 GCGCCGGGCC 428

RESULT 8  
ABI99281  
ID ABI99281 standard; cDNA; 397 BP.  
XX  
AC ABI99281;  
XX  
DT 07-MAR-2002 (first entry)  
XX  
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:123.  
XX  
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.  
XX  
OS Mus musculus.  
XX  
PN WO200188189-A2.  
XX  
XX 22-NOV-2001.  
PD  
XX 18-MAY-2001; 2001WO-JP004192.  
PF  
XX 18-MAY-2000; 2000JP-00145977.  
PR  
XX (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
PA  
XX

PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
XX WPI; 2002-034733/04.  
DR

XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
PT expression levels of particular genes defined in the specification or by  
PT determining the expression profile of a gene group comprising these  
PT genes.  
XX

XX Claim 2; Page 351-352; 2690pp; English.

XX The present invention describes a method for examining ischaemic  
CC conditions, comprising measuring the expression levels of particular  
CC genes (I) in a test sample or determining the expression profile of a  
CC gene group in the sample comprising genes selected from (I). The method  
CC is useful for examining the ischaemic condition (e.g. compressive  
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring the  
CC expression levels of particular genes (ABI99202 to ABI99912, encoding the  
CC protein sequences in AB57020 to AB57374) or by determining the  
CC expression profile of a gene group comprising these genes. The expression  
CC levels or expression profiles produced by these genes are used as an  
CC indicator when screening for ischaemic condition-improving drugs or  
CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR  
CC primers for a mouse ischaemic condition related sequence, which are used  
CC in the exemplification of the present invention  
XX

SQ Sequence 397 BP; 103 A; 105 C; 111 G; 78 T; 0 U; 0 Other;

Query Match 14.3%; Score 241; DB 6; Length 397;  
Best Local Similarity 84.4%; Pred. No. 3.5e-58;  
Matches 271; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1363 ACAGAAACAGGAAAGGTCCTCCGCGAGCTCTGATGCCAAAGTCAAGCTGATCCCAAC 1422  
DB 13 ACAGAAACAGGAAAGGTCCTCCGCGAGCTCTGATGCCAAAGTCAAGCTGATCCCAAC 72  
QY 1423 ACTTTGTGCAACTCCGCGCACTCTATGACCACTGATGATGATGATGATGATGATG 1482  
DB 73 CTTTGTGCACTCCGCGCACTCTATGACCACTGATGATGATGATGATGATGATG 132  
QY 1483 GGAATCTTCAGAACCTGGGCAAGACACCTGCGAGGTGACTCTGGAGGCCCCCTGACC 1542  
DB 133 GGGAACTTCAGAAAGCCCGGATCAGACACCTGCGAGGTGACTCTGGGCGCCCTTAACC 192  
QY 1543 TGTGGAAGGACGGCACCTTACTTACCTCTATGGATAGTGAAGTGGGGCTGGAGTGTGAG 1602  
DB 193 TGTGGAAGGATGGAACTTACTTACCTCTACGGATTTGTAAGCTGGGCGCAGGAATGTGGG 252  
QY 1603 AAGAGCCAGGGGTCTACACCAAGTTACCAAAATTCCTGAATTCGATCAAGCCACCATC 1662  
DB 253 AAGAGCCAGGAGTCTACACTCAAGTCCAAAGTCTCTGAATTCGATTAAGAACCCCATG 312  
QY 1663 AAAAGTGAAAGTGGCTTCTAA 1683  
DB 313 CACAGGAGGCTGGCCTCTGA 333

RESULT 9  
ABA58323/c  
ID ABA58323 standard; DNA; 451 BP.  
XX  
AC ABA58323;  
XX  
DT 01-FEB-2002 (first entry)  
XX  
DE Human foetal liver single exon nucleic acid probe #6628.  
XX  
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157277-A2.  
XX

PD 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US000669.  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human fetal liver.  
XX Claim 1; SEQ ID NO 628; 639pp + Sequence Listing; English.  
XX The invention relates to a single exon nucleic acid probe for measuring  
XX human gene expression in a sample derived from human foetal liver. The  
XX single exon nucleic acid probes may be used for predicting, measuring and  
XX displaying gene expression in samples derived from human fetal liver. The  
XX present sequence is a single exon nucleic acid probe of the invention.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;  
Query Match 7.1%; Score 120.2; DB 4; Length 451;  
Best Local Similarity 97.6%; Pred. No. 1.1e-23;  
Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 324 TCAGAAAGTCAAAATACGTGCAAGGACCAACCATGTGCGCGGGGCAATGTCTCATTAC 383  
Db 451 TCCTACAGTGC AAAATACGTGCAAGGACCAACCATGTGCGCGGGGCAATGTCTCATTAC 392  
QY 384 CCAGAGTCCCTCTACTACCGCTGTGTCTGTAAACACCCCTTACACAGGTCCCGAGTGTCTC 443  
Db 391 CCAGAGTCCCTCTACTACCGCTGTGTCTGTAAACACCCCTTACACAGGTCCCGAGTGTCTC 332  
QY 444 CCAAG 448  
Db 331 CCAAG 327  
RESULT 10  
AAI37947/c  
ID AAI37947 standard; DNA; 451 BP.  
XX  
XX AAI37947;  
XX  
XX 17-OCT-2001 (first entry)  
XX  
XX Probe #6633 used to measure gene expression in human placenta sample.  
XX  
XX Probe; microarray; human; placenta; antenatal diagnosis;  
XX genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157272-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000663.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX

PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488897/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human placenta.  
XX Claim 25; SEQ ID NO 6633; 654pp; English.  
XX The present invention relates to single exon nucleic acid probes (SENP).  
XX The present sequence is one such probe. The probes are useful for  
XX producing a microarray for predicting, measuring and displaying gene  
XX expression in samples derived from human placenta. The probes are useful  
XX for antenatal diagnosis of human genetic disorders  
XX  
SQ Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;  
Query Match 7.1%; Score 120.2; DB 4; Length 451;  
Best Local Similarity 97.6%; Pred. No. 1.1e-23;  
Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 324 TCAGAAAGTCAAAATACGTGCAAGGACCAACCATGTGCGCGGGGCAATGTCTCATTAC 383  
Db 451 TCCTACAGTGC AAAATACGTGCAAGGACCAACCATGTGCGCGGGGCAATGTCTCATTAC 392  
QY 384 CCAGAGTCCCTCTACTACCGCTGTGTCTGTAAACACCCCTTACACAGGTCCCGAGTGTCTC 443  
Db 391 CCAGAGTCCCTCTACTACCGCTGTGTCTGTAAACACCCCTTACACAGGTCCCGAGTGTCTC 332  
QY 444 CCAAG 448  
Db 331 CCAAG 327  
RESULT 11  
AAK32096/c  
ID AAK32096 standard; DNA; 451 BP.  
XX  
XX AC AAK32096;  
XX  
XX DT 06-NOV-2001 (first entry)  
XX  
XX Human bone marrow expressed single exon probe SEQ ID NO: 6653.  
XX  
XX Human; bone marrow expressed exon; gene expression analysis; probe;  
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157276-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000668.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX



CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult liver.  
 CC (i) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis.  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. AB25011-AB51005 represent human  
 CC liver single exon nucleic acid probes of the invention. Note: The  
 CC sequence information for this patent does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;  
 Query Match 7.1%; Score 120.2; DB 4; Length 451;  
 Best Local Similarity 97.6%; Pred. No. 1.1e-23;  
 Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 324 TCAGAAAGTGCAGAAATACGTGCAAGGACACCCATGTGGCGGGCCAAATGTCTCATTTAC 383  
 DB 451 TCCTACAGTGCAGAAATACGTGCAAGGACACCCATGTGGCGGGCCAAATGTCTCATTTAC 392  
 QY 384 CCAGAGTCTCCCTACTACCGCTGTGTCTGTAAACACCCCTTACACAGGTCCAGCTGCTC 443  
 DB 391 CCAGAGTCTCCCTACTACCGCTGTGTCTGTAAACACCCCTTACACAGGTCCAGCTGCTC 332  
 QY 444 CCAAG 448  
 DB 331 CCAAG 327  
 RESULT 14  
 ABS06863/C  
 ID ABS06863 standard; DNA; 451 BP.  
 XX  
 AC ABS06863;  
 XX  
 DT 19-AUG-2002 (first entry)  
 DE Human genome-derived single exon probe from lung SEQ ID No 6854.  
 XX  
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000665.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX

DR WPI; 2002-114183/15.  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples.  
 XX  
 PS Claim 1; SEQ ID NO 6854; 634bp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of probes  
 CC; the novel set of probes which hybridise at high stringency to a nucleic  
 CC acid expressed in the human lung; measuring gene expression in a sample  
 CC derived from human lung, comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of the  
 CC array; identifying exons in a eukaryotic genome, comprising (a)  
 CC algorithmically predicting at least one exon from genomic sequences of  
 CC the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene expression  
 CC analysis, and for identifying exons in a gene, particularly using human  
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 CC present sequence is a single exon probe of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 451 BP; 116 A; 102 C; 127 G; 106 T; 0 U; 0 Other;  
 Query Match 7.1%; Score 120.2; DB 6; Length 451;  
 Best Local Similarity 97.6%; Pred. No. 1.1e-23;  
 Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 324 TCAGAAAGTGCAGAAATACGTGCAAGGACACCCATGTGGCGGGCCAAATGTCTCATTTAC 383  
 DB 451 TCCTACAGTGCAGAAATACGTGCAAGGACACCCATGTGGCGGGCCAAATGTCTCATTTAC 392  
 QY 384 CCAGAGTCTCCCTACTACCGCTGTGTCTGTAAACACCCCTTACACAGGTCCAGCTGCTC 443  
 DB 391 CCAGAGTCTCCCTACTACCGCTGTGTCTGTAAACACCCCTTACACAGGTCCAGCTGCTC 332  
 QY 444 CCAAG 448  
 DB 331 CCAAG 327  
 RESULT 15  
 ABA70919/C  
 ID ABA70919 standard; DNA; 117 BP.  
 XX  
 AC ABA70919;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: May 25, 2004, 05:42:15 ; Search time 3059 Seconds  
(without alignments)  
16429.568 Million cell updates/sec

Title: US-09-912-559-2  
Perfect score: 1683  
Sequence: 1 atgttgccaggatgtctga.....aaagtgaagtggtctcttaa 1683

Scoring table: IDENTITY NUC  
Gapop 10'0 , Gapext 1.0

Searched: 27513289 seqs, 14931030276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_esthum:\*  
4: em\_esthum:\*  
5: em\_estov:\*  
6: em\_estov:\*  
7: em\_estov:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_fod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848.4	50.4	1042	13	EX325239
2	767.4	45.6	1008	13	EX463023
3	698.2	41.5	891	12	BI761782
4	645.4	38.3	889	13	EX431866

C	5	572.8	34.0	576	14	CB156834	CB156834	K-EST0215
	6	533.8	31.7	663	14	CB162144	CB162144	K-EST0222
	7	454.2	27.0	736	14	CB594245	CB594245	AGENCOURT
	8	451.6	26.8	592	10	BE032018	BE032018	330809 WA
	9	447.2	26.6	802	12	BG972579	BG972579	602841155
	10	436.2	25.9	674	9	A3217892	A3217892	mw54B08.r
	11	435.4	25.9	751	12	BI148082	BI148082	602912432
	12	432	25.7	570	12	BM508620	BM508620	4137b03.y
	13	416.4	24.7	929	10	BF788188	BF788188	602113411
	14	399.4	23.7	665	14	CF169114	CF169114	B0808C08-
	15	396.4	23.6	791	12	BI332440	BI332440	602980841
	16	394.2	23.4	546	9	A4237499	A4237499	mx10B10.r
	17	392.6	23.3	608	9	AV601564	AV601564	AV601564
	18	390.4	23.2	902	10	BF780971	BF780971	602105493
	19	386	22.9	682	12	BI220028	BI220028	602934913
	20	384.6	22.9	929	10	BF785781	BF785781	602112402
	21	379.8	22.6	648	14	CF171046	CF171046	B0837E02-
	22	372.4	22.1	969	10	BF384535	BF384535	602046804
	23	370.6	22.0	654	10	AW475402	AW475402	un65h07.y
	24	370.4	22.0	587	12	BG972681	BG972681	602839029
	25	365.6	21.7	659	10	AW610902	AW610902	un13q10.y
	26	363.6	21.6	594	12	EM503097	EM503097	lh51d12.y
	27	363.2	21.6	625	14	CF171125	CF171125	B083BE07-
	28	347.2	20.6	716	10	BF789705	BF789705	602103735
	29	335.8	20.0	655	10	BB569555	BB569555	BB569555
	30	335.4	19.9	668	13	BY742633	BY742633	BY742633
	31	332.6	19.8	732	14	CB599445	CB599445	AGENCOURT
	32	330.8	19.7	934	12	BI765113	BI765113	603051314
	33	327.8	19.5	345	14	T68666	T68666	yc43e09.r1
	34	323	19.2	787	14	CB955374	CB955374	AGENCOURT
	35	317.4	18.9	486	9	AA268125	AA268125	vb08g12.r
	36	311.8	18.5	600	10	BF788541	BF788541	602114406
	37	302.4	18.0	335	14	R89458	R89458	yq01a10.r1
	38	293.2	17.4	104	14	T77362	T77362	yd7g02.r1
	39	286.2	17.0	610	14	CA944054	CA944054	lm05e08.y
	40	279.8	16.6	521	12	BI221626	BI221626	602936942
	41	275.8	16.4	912	12	BG972626	BG972626	602836922
	42	274	16.3	366	14	T93666	T93666	ye06a11.r1
	43	264.4	15.7	434	14	T84369	T84369	yd37g04.r1
	44	258.6	15.4	781	10	BF533788	BF533788	602075308
	45	256.2	15.2	458	14	R10295	R10295	yf36e12.r1

ALIGNMENTS

RESULT 1  
BX325239 1042 bp mRNA linear EST 01-MAY-2003  
LOCUS BX325239 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED  
DEFINITION Homo sapiens cDNA clone CS0DJ010YL24 5-PRIME, mRNA sequence.  
ACCESSION BX325239  
VERSION BX325239.1 GI:30309195  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1042)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 5634.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0AJ010DF12QP1  
&cluster=5634.f. Contact : Feng Liang Email : fliang@lifetech.com  
URL : http://fulllength.invitrogen.com/ invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0AJ010DF12QP1.



```
Db 350 ACCCTGTGAACAGCGTGGGACGTGCTGCTCATGGGAAACCTTCACATCAGCTGCC 409
Qy 299 TGGCTCTCTTCTCTGGAATAATG-TGAGAAAGTGCAAAATACGTGCAAGCAACCC 356
Db 410 TGGCTCTCTTCTCTGGAATAATGATCARAHARATGCAAAATACGTGCAAGCAACCC 469
Qy 357 ATGTGGCGGGGCAATGT--CTCATACCCAGAGTCCTCCCTACTACCGCTGCTCTGT 414
Db 470 ATGTGGCGGGGCAATGTACATCATATACCCAGAGTCCTCCCTACTACCGCTGCTCTGT 529
Qy 415 AAACACCTTACACAGTCCAG-CTGCTCCCAAGTGGTCTCTGTATGAGGCCAAACCC 473
Db 530 AAACACCTTACACAGTCCAGTCTCCCAAGTGGTCTCTGTATGAGGCCAAACCC 589
Qy 474 CTGCAGAAATGGGGCTACCTGCTCCGGCATAGCCAGATCCAAAGTTCACCTGTGCTG 533
Db 590 CTGCAGAAATGGGGCTACCTGCTCCGGCATAGCCAGATCCAAAGTTCACCTGTGCTG 649
Qy 534 TCCGACCACTTCAAGGGGAAATCTGTGAATAGTTCGTATGATGCTGTATGTTGGGA 593
Db 650 TCMGAGCAGWTCAAGGGGAAATCTGTGAATAGTTCGTATGATGCTGTATGTTGGGA 709
Qy 594 TGGCTACTCTTACCGAGGAAATGAATAGGACAGTCAACACGATGCGTGCCTTTACTG 653
Db 710 TGGCTACTCTTACCGAGGAAATGAATAGGACAGTCAACACGATGCGTGCCTTTACTG 769
Qy 654 GAATCTCCACTCTCTTGAGGAGAAATTACAAATGTTTATGAGGATGCTGAAACCCA 713
Db 770 GAATCTCCACTCTCTTGAGGAGAAATTACAAATGTTTATGAGGATGCTGAAACCCA 829
Qy 714 TGGGATGGGGAAACACAAATTTCTGCAGAAACCCAGATGGGACGAAAGCCCTGG-TGCT 772
Db 830 TGGGATGGGGAAACACAAATTTCTGCAGAAACCCAGATGGGAGAAAGCCCTGTATGCT 889
Qy 773 TTATTAAGTTACCAATGACAGGTGAAT-TGGGAATCTGTATGCTCAGCCTGCTCA 831
Db 890 TTATTAAGTTACCAATGACAGGTGAAT-TGGGAATCTGTATGCTCAGCCTGCTCA 949
Qy 832 GCCCAGGACCTTGCTTACCAGAGGAAAGCCCACTGAGCCATCAACCAAGTTCGCGG 890
Db 950 ACCASRACGTGCTAMCCAGCGAGAGCCCACTGAGAGCATCAACCAARCTTCGCGG 1008

RESULT 3
BI761782
LOCUS 603046775f1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5187066 5',
DEFINITION mRNA sequence.
ACCESSION BI761782
VERSION BI761782.1 GI:15753360
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: csapps-x@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M11467 row: e column: 19
High quality sequence stop: 839.
Location/Qualifiers
1..891
```

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5187066"
/lab_host="DH10B"
/clone_lib="NIH_MGC_116"
/note="Organ: pooled colon, kidney, stomach; Vector:
PCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous. Pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon, 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."
```

## ORIGIN

```
Query Match 41.5%; Score 698.2; DS 12; Length 891;
Best Local Similarity 97.2%; Pred. No. 4.5e-188;
Matches 775; Conservative 0; Mismatches 13; Indels 9; Gaps 6;

Qy 1 ATGTTTGGCAGAGTGTCTGATCTCCATGTTCTGCTGTTTAAATGGCTCTGTTGGAAAGACA 60
Db 81 ATGTTTGGCAGAGTGTCTGATCTCCATGTTCTGCTGTTTAAATGGCTCTGTTGGAAAGACA 140
Qy 61 GCCTGTGGGTTCTCCCTGATGTTCTTATTTGGAAGCCCTGGACCCAGACTGGACCCCTGAC 120
Db 141 GCCTGTGGGTTCTCCCTGATGTTCTTATTTGGAAGCCCTGGACCCAGACTGGACCCCTGAC 200
Qy 121 CAGTATGATTACAGCTTACGAGGATTTAATCAGGAGAGAGACACACAGTAGCACATTACC 180
Db 201 CAGTATGATTACAGCTTACGAGGATTTAATCAGGAGAGAGACACACAGTAGCACATTACC 260
Qy 181 CATCTGAGAAATCTTGACTGTGTACTACCTGAGGACCAAGCTGATTCATGCGAGCCCAAC 240
Db 261 CAGCTGTAGAAATCTTGACTGTGTACTACCTGAGGACCAAGCTGATTCATGCGAGCCCAAC 320
Qy 241 CCCTGTGAACAGGTGGGACTGCTCTGTCATGAGGACCTTACATGAGCTGCTGCTG 300
Db 321 CCCTGTGAACAGGTGGGACTGCTCTGTCATGAGGACCTTACATGAGCTGCTGCTG 380
Qy 301 GCTCTCTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTGGAAGGACAAACCATGT 360
Db 381 GCTCTCTTCTCTGGGAATAAGTGTGAGAAAGTGCAGAAATACGTGGAAGGACAAACCATGT 440
Qy 361 GGCCGGGGCAATGCTCATTTACCCAGAGTCTCCCTACTACTACCGCTGTGTCTGTAACAC 420
Db 441 GGCCGGGGCAATGCTCATTTACCCAGAGTCTCCCTACTACTACCGCTGTGTCTGTAACAC 500
Qy 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGTTCCTGTATGAGGACCAACCCCTGCCAG 480
Db 501 CTTTACAGAGTCCAGCTGCTCCCAAGTGTTCCTGTATGAGGACCAACCCCTGCCAG 560
Qy 481 AATGGGGCTACCTGCTCCCGCATAGCGGAGATCCAAAGTTCACCTGTGCTGCTCCGAC 540
Db 561 AATGGGGCTACCTGCTCCCGCATAGCGGAGATCCAAAGTTCACCTGTGCTGCTCCGAC 620
Qy 541 CAGTTCAAGGGGAAATTTCTGTGAAAATAGTGTCTGATGATGCTGATGTTGGGATGGCTAC 600
Db 621 CAGTTCAAGGGGAAATTTCTGTGAAAATAGTGTCTGATGATGCTGATGTTGGGATGGCTAC 680
Qy 601 TCTTACCGAGGAAATTAATAGGACAGTC-AACAGCATGCGTGCCTTTACTGGAATC 659
Db 681 TCTTACCGAGGAAATTAATAGGACAGTC-AACAGCATGCGTGCCTTTACTGGAATC 740
Qy 660 CC-ACCTCTCTTTCAGGAGAAATTAACATGTTTATGGAGG---ATGCTGAACCCCATG 715
Db 741 CCAAGCTCTCTTTCAGGAGAAATTAACATGTTTATGGAGGATGCTTGAACCCCATG 800
Qy 716 GGATTGGGAA--CACAAATTTCTGCAG-AAACCCAGATGCGGACG-AAACCCCTGCTGTC 771
```



Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail  
Plate: 17 row: H C  
High quality sequence

FEATURES	SOURCE
1. <b>Feature 1</b>	Source 1
2. <b>Feature 2</b>	Source 2
3. <b>Feature 3</b>	Source 3
4. <b>Feature 4</b>	Source 4
5. <b>Feature 5</b>	Source 5
6. <b>Feature 6</b>	Source 6
7. <b>Feature 7</b>	Source 7
8. <b>Feature 8</b>	Source 8
9. <b>Feature 9</b>	Source 9
10. <b>Feature 10</b>	Source 10
11. <b>Feature 11</b>	Source 11
12. <b>Feature 12</b>	Source 12
13. <b>Feature 13</b>	Source 13
14. <b>Feature 14</b>	Source 14
15. <b>Feature 15</b>	Source 15
16. <b>Feature 16</b>	Source 16
17. <b>Feature 17</b>	Source 17
18. <b>Feature 18</b>	Source 18
19. <b>Feature 19</b>	Source 19
20. <b>Feature 20</b>	Source 20
21. <b>Feature 21</b>	Source 21
22. <b>Feature 22</b>	Source 22
23. <b>Feature 23</b>	Source 23
24. <b>Feature 24</b>	Source 24
25. <b>Feature 25</b>	Source 25
26. <b>Feature 26</b>	Source 26
27. <b>Feature 27</b>	Source 27
28. <b>Feature 28</b>	Source 28
29. <b>Feature 29</b>	Source 29
30. <b>Feature 30</b>	Source 30
31. <b>Feature 31</b>	Source 31
32. <b>Feature 32</b>	Source 32
33. <b>Feature 33</b>	Source 33
34. <b>Feature 34</b>	Source 34
35. <b>Feature 35</b>	Source 35
36. <b>Feature 36</b>	Source 36
37. <b>Feature 37</b>	Source 37
38. <b>Feature 38</b>	Source 38
39. <b>Feature 39</b>	Source 39
40. <b>Feature 40</b>	Source 40
41. <b>Feature 41</b>	Source 41
42. <b>Feature 42</b>	Source 42
43. <b>Feature 43</b>	Source 43
44. <b>Feature 44</b>	Source 44
45. <b>Feature 45</b>	Source 45
46. <b>Feature 46</b>	Source 46
47. <b>Feature 47</b>	Source 47
48. <b>Feature 48</b>	Source 48
49. <b>Feature 49</b>	Source 49
50. <b>Feature 50</b>	Source 50
51. <b>Feature 51</b>	Source 51
52. <b>Feature 52</b>	Source 52
53. <b>Feature 53</b>	Source 53
54. <b>Feature 54</b>	Source 54
55. <b>Feature 55</b>	Source 55
56. <b>Feature 56</b>	Source 56
57. <b>Feature 57</b>	Source 57
58. <b>Feature 58</b>	Source 58
59. <b>Feature 59</b>	Source 59
60. <b>Feature 60</b>	Source 60
61. <b>Feature 61</b>	Source 61
62. <b>Feature 62</b>	Source 62
63. <b>Feature 63</b>	Source 63
64. <b>Feature 64</b>	Source 64
65. <b>Feature 65</b>	Source 65
66. <b>Feature 66</b>	Source 66
67. <b>Feature 67</b>	Source 67
68. <b>Feature 68</b>	Source 68
69. <b>Feature 69</b>	Source 69
70. <b>Feature 70</b>	Source 70
71. <b>Feature 71</b>	Source 71
72. <b>Feature 72</b>	Source 72
73. <b>Feature 73</b>	Source 73
74. <b>Feature 74</b>	Source 74
75. <b>Feature 75</b>	Source 75
76. <b>Feature 76</b>	Source 76
77. <b>Feature 77</b>	Source 77
78. <b>Feature 78</b>	Source 78
79. <b>Feature 79</b>	Source 79
80. <b>Feature 80</b>	Source 80
81. <b>Feature 81</b>	Source 81
82. <b>Feature 82</b>	Source 82
83. <b>Feature 83</b>	Source 83
84. <b>Feature 84</b>	Source 84
85. <b>Feature 85</b>	Source 85
86. <b>Feature 86</b>	Source 86
87. <b>Feature 87</b>	Source 87
88. <b>Feature 88</b>	Source 88
89. <b>Feature 89</b>	Source 89
90. <b>Feature 90</b>	Source 90
91. <b>Feature 91</b>	Source 91
92. <b>Feature 92</b>	Source 92
93. <b>Feature 93</b>	Source 93
94. <b>Feature 94</b>	Source 94
95. <b>Feature 95</b>	Source 95
96. <b>Feature 96</b>	Source 96
97. <b>Feature 97</b>	Source 97
98. <b>Feature 98</b>	Source 98
99. <b>Feature 99</b>	Source 99
100. <b>Feature 100</b>	Source 100

```
1. .663
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L17N670205n1-17-H10"
sex="F"
```

```

clone="L17N670205n1-17-H10"
sex="F"
lab host="Top10F."
clone lib="L17N670205n1"
note="Organ: Liver; Vector: pT73-Pac; Site: 1: ECRI; Site 2: NCI; The library was contributed by The Soares Laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from harvested cell culture."

```

Crone, M.D. [mdcrone@uic.edu](mailto:mdcrone@uic.edu)  
 Crone Labs, University of Illinois at Chicago  
 1601 S. Morgan St., Room 608, Chicago, IL 60607-7136  
 Site 2: NotI. The library was contributed by the Soares  
 Laboratory and it was constructed as described by Bonaldo,  
 M.F.F., Lennon, G. and Soares, M.B. (1996), Genome Research  
 6(9): 791-806. RNA was prepared from harvested cell  
 culture."

M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 66(9): 791-806. RNA was prepared from harvested cell culture."

31.7%; Score 533.8; DB 14; Length 663;

TTTCATCAGCAGAGCTTTAGGGTGCAAGATATTCAGTACAGCCACTACAAT 1206

TTTCATGAGCAGAGCTTTAGGGTGGAGATAGATATTCAGGTACAGGCCATACATTT 99

ATGAGATTCCCCACAATGATATTGCATTGCTCAAGTTAAAGCCAGTGGATGGT 120

CTCTAGAATCCAAATACGTGAGACTGTGTGCTTGCCTGATGGGTCTTTTCCC 180

GTGAGTGCCACATCTCTGGCTGGGTGTTACAGAAACAGGAAAGGGTCCCGC 240

TTGGATGCCAAGTCAAGCTGATTGCCAACACTTTGTGCAACTCCCGCCAACTC 300

ACATCATTTGATGCAGTATGATCTGTGCAGGAATCTTCAGAACTGGGCAA 360

[illegible]

GGGATAGTGAGCTGGGGCCCTGGAGTGTGAGGAGGCCAGGGGTCTACACCCAA 1620

AAATTCCTGAATTGGATCAAGCCACCATCAAAAGTGAAAGTGGCTTCTAA 1683

	NIH MGC 177	Mus musculus	cDNA clone	mRNA	linear	EST 03-APR-2003
	12929103			736 bp		

1 GR:29512101

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 736)

NIH-MGC <http://mgi.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)

Contact: Robert Strausberg, Ph.D.  
Email: [csabp-remail.nih.gov](mailto:csabp-remail.nih.gov)

Tissue Procurement: Dr. Michael Brownstein  
cDNA Library Preparation: Michael Brownstein Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: NDCW92 row: d column: 23  
High quality sequence stop: 558.

Location/Qualifiers

1. 736

source

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:30310462"

/lab\_host="DH10B (TI-phage-resistant)"

/clone\_lib="NIH MGC 177"

/notes="Organ: liver; Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggcgccctggcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:  
5'-AAGCAGGTGATCAGCAGGATGCGCATACGCGCGG-3' and  
5'-ATTCTAGCGCCAGCGCGGCATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 27.0%; Score 454.2; DB 14; Length 736;  
Best Local Similarity 80.8%; Pred. No. 2.9e-118;  
Matches 542; Conservative 0; Mismatches 128; Indels 1; Gaps 1;

QY 715 GGGATTGGGACACAACTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGGTCTTT 774  
DB 3 GGGATCCAGAGACAACTTCTGAGAAACCCAGATGCGGACGAAAGCCCTGGTCTTT 52  
QY 775 ATTAAGTTACCAATGACAAGGTGAATGGGAATCTGTGATCTCAGCCTGCTCAGCC 834  
DB 63 GTCAAGGTGAACAGTGAAGAGTGAATGGGAATCTGTGATCTCAGCCTGCTCAGTG 122  
QY 835 CAGGACCTTGCCTACCCAGAGGAGCCCACTGAGCCATCAACCAAGCTTCGGGGTTT 894  
DB 123 CTTGACACCCCTAACCCAGTGAAGCCCTTCTGGAGCCTGTGTGAGCTGCCAGGTTTC 182  
QY 895 GACTCTCTGTGGAAGACTGAGATAGCAGAGGAGGAGATCAAGAGAATCTATGAGGCTTT 954  
DB 183 GAGTCTCTGGGAGAGCGAGGTAGTGAACAGCAGTCAAGCGTATCTACGGGGCTTT 242  
QY 955 AAGAGCAGGGGGGCAAGCACCATGCGAGGCTCCCTCCAGTCCCTGCTGCTCTGACC 1014  
DB 243 AAGAGCAGAGGAGCAAGCACCCTGCGAGGTGTCCTGAGACCTTCACTGCGGTTGACC 302  
QY 1015 ATCTCCATGCCCCAGGCGCACTTCTGTGTGGGGGCTGTATCCACCCCTCTGGGTGCTC 1074  
DB 303 ACCTCCATGCCCCAAGGCCACTTCTGTGGGGGGCCCTGTATCCACCCCTCTGGGTGCTC 362  
QY 1075 ACTGCTGCCACTCTACCGACATATAAACACAGACATCTAAGGTGGTGTAGGGACACG 1134  
DB 363 ACTGACCCCACTGTACCGACATATAAACACAGACATCTAAGGTGGTGTAGGGACACG 422  
QY 1135 GACCTGAAGAAGAAGATTTCATGAGCAGAGCTTTAGGGTGCAGAGATATTCAAGTAC 1194  
DB 423 GACCTGAAGAAGACAGATCCCATGACAGACCTTCAGGGTGGGAAAAAATCTGAAGTAC 482

QY 1195 AGGCACATCAATGAAGAGATGAGATTCGCCCAATGATATTCATTGCTCAAGTTAAAG 1254  
DB 483 AGTCAGTATATGAAGAGATGAGATTCGCCCAATGATATTCATTGCTCAAGTTAAAG 542

QY 1255 CCAGTGTATGCTCACTGTGCTCTAGAAATCCAAATAGCTGAAGACTGTGCTTTCCTGAT 1314  
DB 543 CCAGTGTGCTCACTGTGCTCTGGAATCCAGATATGAGACTGTATGTTGCCAGC 602

QY 1315 GGGTCTTTTCCCTCTGGAAGTGGAGTCCCATCTCTGGCT-GGGTGTACAGAAACAG 1373  
DB 603 GACCCCTTTCCCTCTGGAAGTGGAGTCCCATCTCTGGCTGGGGGTGTTTCAGAAACAG 662

QY 1374 AAAAGGGTCCC 1384  
DB 663 GGAAGGGGTCC 673

RESULT 8  
BE032018

LOCUS BE032018 592 bp mRNA linear EST 09-JUL-2000

DEFINITION 130809 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.

ACCESSION BE032018

VERSION BE032018.1 GI:8327027

KEYWORDS EST.

SOURCE Sus scrofa (pig)

ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 592)  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
Fahrenkrug, S.C., Smith, T.P.L., Freking, B.A., Cho, J., White, J., Quackenbush, J., and Keele, J.W.  
Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly  
Mamm. Genome 13 (8), 475-478 (2002)

JOURNAL 22213789  
MEDLINE 12226715  
PubMed

COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: [smithemail.marc.usda.gov](mailto:smithemail.marc.usda.gov)  
Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGTCACGACG  
Plate: 69 row: K column: 12  
Seq primer: ATTTAGGTGACACTATAG.  
Location/Qualifiers  
1. 592  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="MARC 1P1G"  
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
Library made from pooled tissue from day 11, 13, 15, 20,  
and 30 embryos."

FEATURES

source

ORIGIN

Query Match 26.8%; Score 451.6; DB 10; Length 592;  
Best Local Similarity 86.0%; Pred. No. 1.4e-117;  
Matches 518; Conservative 0; Mismatches 69; Indels 15; Gaps 1;

QY 522 CACCTGTGCTGTCGCCAGCTTCAAGGGGAAATCTGTGAATAGCTTCTGATGACTG 581  
DB 6 CGCGTGGGCTGCTCCTGACCACTTAAAGGGGAGATTTCTGGAATAGGTTCTGATGACTG 65

QY 592 CTATCTGGCGTGTCTCTTCCGAGGGGAAATGAATAGGACAGTCAACACGATGC 641





AA217892.1 GI:1826875  
Mus musculus (house mouse)  
EST.  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Possible reversed clone: similarity on wrong strand  
Seq primer: -28ml3 rev2 BT from Amersham  
High quality sequence stop: 485.  
Location/Qualifiers  
1. 674  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:658839"  
/sex="unknown"  
/tissue\_type="fetus"  
/dev\_host="12.5dpc total fetus"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse 3NME12 5"  
/note="Organ: whole fetus; Vector: pTV73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACATCTGAGTGGAGCGCGCGCTTATTTTATTTTATTTT 3'], on total mouse RNA [provided by Minoru Ko, Wayne State Univ.]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTV73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

Db 241 TCTACGGGGCTTTAAGAGCAGCAGGCAAGCAGCCCGTGGCAGGTGTCTCCCTGCAGACCT 300  
Qy 1001 CGCTGCTCTGACCATCTCATGCCCCAGGGCCACTTCTGTGTGGGGCGCTGATCCACC 1060  
Db 301 CACTGCGGTTGACCACTCCATGCCCAAGGCCACTTCTGTGGGGCGCCCTGATCCACC 360  
Qy 1061 CCTGCTGGGTGCTCACTGTGCCCACTGCACCGACATAAAACCAACACATCTAAAGGTG 1120  
Db 361 CCTGCTGGGTGCTCACTGTGCCCACTGTACCGACATAAACCAACACATCTAAAGGTG 420  
Qy 1121 TGTAGGGGACCAGGACCTGGAAGAAAGAAATTTATGAGCAGAGCTTTAGGCTCAGA 1180  
Db 421 TACTAGGGGATCAGGACCTGGAAGAAAGACAGATCCCATGACAGACCTTCAGGGTGA 480  
Qy 1181 AGATATTCAGTACAGCCACTCAATGAAGAGATGAGATCCCAATGATATTCAT 1240  
Db 481 AAATACCTGAAGTACAGTCACTATATGAAGAGATGAGATTTCCCAATGATATTCCT 540  
Qy 1241 TGTCTCAAGTTAAAGCCAGTGGTCACTGTCTCTAGAAATCCAAATACGTGAAGACTG 1300  
Db 541 TGTCTCAAGTTAAAGCCAGTGGTCACTGTCTCTAGAAATCCAGATATGTGAAGACTG 600  
Qy 1301 TGTCTCTGCTGATGGGTCTTCCCTCTGGAGTGTAGTCCACATCTCTGGCTGGGGTG 1360  
Db 601 TATGTTTGCACAGGACCCCTTTGCTTGGACTGAGTGCACATCTCTGGACTGGGTG 660  
Qy 1361 TTACAGAAACAGG 1373  
Db 661 GTACAGAAACAGG 673

RESULT 11  
BI148082  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BI148082 751 bp mRNA linear EST 05-JUL-2001  
602912432F1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:5053663 5',  
mRNA sequence.  
BI148082.1 GI:14608083  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 751)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Inyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM1145 row: 0 column: 08  
High quality sequence stop: 747.  
Location/Qualifiers  
1. 751  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5053663"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI CGAP L19"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

FEATURES  
source  
ORIGIN

AA217892.1 GI:1826875  
Mus musculus (house mouse)  
EST.  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Possible reversed clone: similarity on wrong strand  
Seq primer: -28ml3 rev2 BT from Amersham  
High quality sequence stop: 485.  
Location/Qualifiers  
1. 674  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:658839"  
/sex="unknown"  
/tissue\_type="fetus"  
/dev\_host="12.5dpc total fetus"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse 3NME12 5"  
/note="Organ: whole fetus; Vector: pTV73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACATCTGAGTGGAGCGCGCGCTTATTTTATTTTATTTT 3'], on total mouse RNA [provided by Minoru Ko, Wayne State Univ.]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTV73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

Query Match 25.9%; Score 436.2; DB 9; Length 674;  
Best Local Similarity 78.0%; Pred. No. 4e-113;  
Matches 525; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

Qy 701 ATGCTGAACCCATGGGATGGGCAACACATTTCTCAGAAACCCAGATGGCGACGAAA 760  
Db 1 ATTGGGAAGAGCGGATGCGAGACACATCTCTGAGNAACCCAGATGGAGACACA 60  
Qy 761 AGCCCTGTGCTTTATTAAGTTACCAATGACAGGTGAATGGGAATACCTGTATGCT 820  
Db 61 AACCTGTGTTTCTGCTCAAGGTGACAGTGAAGAGTGAATGGGAATACCTGTATGCTCA 120  
Qy 821 CAGCTGTCTAGCCAGACAGCTTGCTTACCCAGAGGAAGCCCACTGAGCATCAACCA 880  
Db 121 CAGTGTGCTAGTCTGACACCCCTAACCCAGTGGAAAGCCTTCTGGAGCCTGTGATGG 180  
Qy 881 AGCTTCCGGGGTTTGACTCTCTGTGAAAGACTGAGATAGCAGAGGAGGAAGATCAAGAA 940  
Db 181 AGCTCCAGGGTTTCAGTCTCTGCGGAAGACGGAGTACCTGAACACGACAGTCAAGGTA 240  
Qy 941 TCTATGGAGGCTTTAAGAGCAGCGCGGCGAGACCCCATGGCAGGCTCCCTCCAGTCT 1000



Db 183 TATGTGAAGACTGTATGTTGGCCAGGACCCCTTCCCTCTGGAAGTGAAGTGCACATC 242  
 QY 1348 TGTGGCTGGGGTGTATACAGAAACAGAAAGGGTCCGCCAGCTCCTGGATGCCAAAGTC 1407  
 Db 243 TGTGGCTGGGGTGTATACAGAAACAGGGAAGGGTCCGCCAGCTCCTGGATGCTAAAGTC 302  
 QY 1408 AAGCTGATTGCCAACACATTTGTGCAACTCCGCCAACTCTATGACCAATGATGATGAC 1467  
 Db 303 AAGCTAATCGTAAACCTTTGTGCAACTCCGCCAACTCTATGACCAATGATGATGAC 362  
 QY 1468 AGTATGATCTGTGAGGAAATCTTACAGAAACCTGGGCAAGACACTGCGCAGGGTGAATCT 1527  
 Db 363 AGTATGATTTGTGCGGGAACTTTCAGAAAGCCGAGTACAGACACTGCGCAGGGTGAATCT 422  
 QY 1528 GGAGGGCCCTGACCTGTGAGAAAGACCGCACTACTACTAGCTCTATGGATAGTGAAGTGG 1587  
 Db 423 GGGGGCCCTCTAAGCTGTGAGAGGATGGAATCTACTAGCTCTACGGATGTAAGTGG 482  
 QY 1588 GGCCTGGAGTGTAGAGAGGCGAGGGTCTTACAGAACTTACAGAACTTCAAAATCTGTAATGG 1647  
 Db 483 GGCAGGAATGTGGGAAGAGCAGGAGTCTACAGTCAAGTCAAGTCAAGTCTCTGTAATGG 542  
 QY 1648 ATCAAGGCCACCATCAAAAGTGAAGTG 1675  
 Db 543 ATAAAGACCACCATGCACAGGAGGCTG 570

## RESULT 13

BF788188 929 bp mRNA linear EST 12-JAN-2001  
 LOCUS 602113411F1 NCI\_CGAP\_Kid14 Mus musculus cDNA clone IMAGE:4241642  
 DEFINITION 5', mRNA sequence.

ACCESSION BF788188

VERSION BF788188

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LIA9858 row: e column: 03

High quality sequence stop: 686.

Location/Qualifiers

1..929

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:4241642"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_lib="NCI CGAP Kid14"

/note="Organ: kidney; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.75 kb. Constructed by Life

Technologies. Note: this is a NCI\_CGAP Library. |"

## ORIGIN

Query Match 24.7%; Score 416.4; DB 10; Length 929;  
 Best Local Similarity 75.9%; Pred. No. 2.3e-107;  
 Matches 529; Conservative 0; Mismatches 162; Indels 6; Gaps 1;

QY 1 AAGTTGCCAGAGTGTCTGATCTCCATGTTCTGCTGTTAATGCTCTGTTGGGAAAGACA 60  
 Db 44 ATATTGTGACAGATGTAGGTGTTCCGTGCTCTGCTGCTATATCGCCCTGTTGGGAAAGTCA 103  
 QY 61 GCCTGTGGGTCTCCCTGATGCTTTATTGGAAAGCTGACCCAGACTGGACCCCTGAC 120  
 Db 104 GTCATGGGGCTCTCACTGATGCTCTTCATTGCGCCCCCAGACCCAGATGGACCCCCGAT 163  
 QY 121 CAGTATGTTACAGCTACGAGGATTAATCAGGAAGAGAACACCACTAGTAGCACACTTACC 180  
 Db 164 GACTATTACTACAGCTATGAGCAGTCCAGCCCAAGAGACCCCACTGTCACGAGAC 223  
 QY 181 CATGCTGAGAACTCTGATGCTGCTACTACACTGAGGACCAAGCTGATCCATGCGACCCCAAC 240  
 Db 224 ACCCTGAGAACCCCGACTGCTACT-----ATGAAGACGATGATCCATGCGACCTCCAAC 277  
 QY 241 CCCTGTGAACCGTGGGACTGCTCTGCTCATGGGAGCACCTTTCATGCACTGACCTGCTG 300  
 Db 278 CCCTGTGAACCGCGGGGACTGCTATCATCAGAGGGGATACCTTTCAGTTGCACTGCCCCA 337  
 QY 301 GCTCCTTTCTCTGGGAATAAGTGTTCAGAAAGTGCAGAAATACGTGCAAGGACCAACCCATGT 360  
 Db 338 GCCCCTTTCTCGGGAGCCGCTGCCAGACTGCACAAAACAAGTGCAGGACCAACCCATGT 397  
 QY 361 GCGCGGGCCAAATGCTCATTAACCCAGAGTCTCCCTACTACCGCTGTGCTGTAAACAC 420  
 Db 398 GTCCATGTGTAATGCTCTCATTTACCCAGAGACCCCTACTACCGCTGTGCTGCAATATC 457  
 QY 421 CTTACACAGTCCCACTGCTCCCAAGTGTCTCTGTATGTCAGGCAAAACCCCTGCCAG 480  
 Db 459 CTTACACGGACAGACTGCTCCAAAGTCTTCGGCATGCAGGCCAAACCCCTGCCAG 517  
 QY 481 AATGGGCTACTCTCCCGCATACCGGAGATCCAGTTACCTGCTGCTGCTGCTGCTGCTG 540  
 Db 518 AATGGGAGTCTGTTCCCGACACAGCGAGATCCAGTTTACCTGCTGCTGCTGCTGCTG 577  
 QY 541 CAGTTCAAGGGGAAATCTCTGAAATAGTTCCTGATGACTGCTATGTTGGCGATGCTAC 600  
 Db 578 CAGTATAGGGGAAATCTCTGAAATAGTTCCTGAAATAGTTCCTGAAATAGTTCCTG 637  
 QY 601 TCTTACCGAGGAAATGTAATAGACAGTCAACAGCATGCGTGCCTTACTTACTGGAATCT 660  
 Db 638 TCTTACCGAGGAAATGTAATAGACAGTCAACAGCATGCGTGCCTTACTTACTGGAATCT 697  
 QY 661 CACTCTCTTTCGAGGAGAAATTAACAATGTTATGG 697  
 Db 698 CACTCTCTTTCGAGGAGAAATTAACAATGTTATGG 734

## RESULT 14

CF169114

LOCUS CF169114

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

CF169114 665 bp mRNA linear EST 25-JUL-2003  
 E0809C08-5 NIA Mouse Newborn Kidney cDNA Library (Long 1) Mus  
 musculus cDNA clone NIA:E0809C08 IMAGE:30468319 5', mRNA sequence.  
 CF169114  
 CF169114.1 GI:33278663  
 EST.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 665)  
 Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.  
 Construction of long-transcript enriched cDNA libraries from  
 submicron amounts of total RNAs by a universal PCR amplification  
 method  
 Genome Res. 11 (9), 1553-1558 (2001)  
 21429098  
 11544199  
 Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@lgsun.grc.nia.nih.gov  
Plate: B0809 row: C column: 08  
Seq primer: M13 Reverse  
High quality sequence stop: 665  
POLYA=No.

Location/Qualifiers  
1. .665  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="nia:EST:B0809C08-5"  
/clone="NIA:B0809C08 IMAGE:3046819"  
/dev\_stage="Newborn Kidney"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse Newborn Kidney cDNA Library (Long 1)"  
/note="vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). In brief, double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen]; 5'-TGACTAGTTCGAGCGCGCCGCTTTT-3' from 26 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol precipitation. The cDNAs were ligated to lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0 kb. The library was constructed by Yulan Piao."

FEATURES  
source

## ORIGIN

Query Match 23.7%; Score 399.4; DB 14; Length 665;  
Best Local Similarity 76.8%; Pred. No. 1.4e-102;  
Matches 503; Conservative 0; Mismatches 146; Indels 6; Gaps 1;

Qy 1 ATGTTTGCCAGGATCTGATCTCCATGTTCTGCTGTTAATGGTCTGGTGGGAAGACA 60  
Db 17 ATATTTGTCAGGATGTTGGTGTTCGTTCTGCTGCTAATCGCCCTGGTGGGAAGTCA 76  
Qy 61 GCCTGTGGGTTCTCCCTGATGTCCTTTATTGGAAGCCCTGGACCCAGACTGACCCCTGAC 120  
Db 77 GTCAATGGGCTCTCACTGATGTCCTTCATTGGCGCCCCAGACCCAGATTGGACCCCGAT 136  
Qy 121 CAGTATGATTACAGCTACGAGGATTATATCAGGAAGAGAACACCAAGTAGCACACTTACC 180  
Db 137 GACTATTACTAGCTATGAGCAGTCCAGCCAGAGAGACCCAGTGTCAAGCAGACC 196  
Qy 181 CATGTGAGATCTCTGATGCTGCTACACTGAGGACCAAGTGTATCCATGCGAGCCCAAC 240  
Db 197 ACCCTGTGAGACCCGAGCTGGTACT-----ATGAAGACGATGATCCATGCGAGTCCAAC 250  
Qy 241 CCCTGTGAACAGGGTGGGAGCTGCTCGTCCATGGGAGACCTTCACATGAGCTGCGCTG 300  
Db 251 CCCTGTGAACAGGGGAGCTGTTATCATCAGAGGGGATACCTTCAGTTGAGCTGCCCA 310  
Qy 301 GCTCCTTTCTCTGGGAATAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAAACCCATGT 360  
Db 311 GCCCCCTTCTCGGGAGCGGTGCAGACTGCACAAACAAAGTGCAGGACAAACCCATGT 370  
Qy 361 GGCGGGGCCATGTCATTAACCCAGAGTCTCCCTACTACCGCTGTCTGTAACAC 420  
Db 371 GTCCATGTGTGCTGCTATTACCCAGAGACCCCTACTACCGCTGTGCTGCAATAC 430  
Qy 421 CTTTACACAGGTCCTGCTCCCAAGTGTCTCTGTATGAGGCCAAACCCCTGCCAG 480

Db 431 CCTTACAGGGACCAGACTGCTCCAAAGTGTTCGGCATCGAGGCCAAACCCCTGCCAG 490  
Qy 481 AATGGGGTACTCTGCTCCCGGATAGCGGAGATCCAAAGTTACCTGTGCTGTCCCGAC 540  
Db 491 AATGGCGAGTCTGTTCCGACACAGAGGAGATCCAGGTTACTGTGCTGTCCAGAC 550  
Qy 541 CAGTTCAAGGGGAATCTGTGAATAGTGTCTGATGATCTGCTATGTGGGATGGCTAC 600  
Db 551 CAGTATAAGGGGAATCTGTGAATAGTGTCCGGACGACTGTATGTGCTGCTGATGGCTAC 610  
Qy 601 TCTTACCGGGGAATGAATAGGACAGTCAACACGATCGTGTGCTTACTTGGGA 655  
Db 611 TCTTACCGGGGAATGAGTAGACAGTCAACACGATCGTGTGCTTACTTGGGA 665

RESULT 15  
LOCUS B1332440  
DEFINITION B1332440 791 bp mRNA linear EST 30-JUL-2001  
602380841F1 NCI\_CGAP\_Li9 Mus musculus cDNA clone IMAGE:5133587 5',  
mRNA sequence.  
ACCESSION B1332440  
VERSION B1332440.1 GI:15017097  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 791)  
NIH-MGC http://mgc.ncl.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-rc@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
CDNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLML1328 row: a column: 12  
High quality sequence stop: 787.

FEATURES  
source

1. .791  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5133587"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="NCI\_CGAP\_Li9"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN

Query Match 23.6%; Score 396.4; DB 12; Length 791;  
Best Local Similarity 75.6%; Pred. No. 1.1e-101;  
Matches 520; Conservative 0; Mismatches 161; Indels 7; Gaps 2;

Qy 1 ATCTTCCAGGATGCTGATCTCCATGTTCTGCTGTTAATGGCTCTGTTGGGAAGACA 60  
Db 23 ATATTTGTACGAGTGTGGTGTTCCTGTCTCTGCTAATCGCCCTGTGTTGGGAATCA 82  
Qy 61 GCCTGTGGGTTCTCCCTGATGTCCTTTATTGGAAGCCTCGACCCAGACTGGACCCCTGAC 120  
Db 83 GTCATTTGGGCTCTCACTGATGTCCTTCATTGGCCCCCAGACCCAGATTGGACCCCGAT 142  
Qy 121 CAGTATGATTACAGCTACGAGGATTATATCAGRAGAGACACCACTAGTACACTTACC 180  
Db 143 GACTATTACTACGCTATGAGCAGTCCAGCCAGAGACCCAGTGTCAAGCAGACC 202

QY	181	CATGCTGAGATCCTGACTGTACTACACTGAGGACCAAGCTGATCCATGCCAGCCCAAC	240
Db	203	ACCCCTGAGAACCCCGACTGGTACT------ATGAAGACGATGATCCATGCCAGTCCCAAC	256
QY	241	CCCTGTGAACACGGTGGGACTGCTCCTGTCATGGGAGCAGCTTCACATGACAGTGGCTG	300
Db	257	CCCTGTGAACACGGGCGGACTGTATCATCAGAGGGGATACCTTCAGTTGACAGTGGCCCA	316
QY	301	GCTCCTTCTCTGGGAATAGTGTGAGAAAGTGCAAAATAGTGCAGAGGACAAACCCATGT	360
Db	317	GCCCCCTTCTCGGGGAGCGGGTGCAGAGCTGCACAAAACAAGTGCAGGACAAACCCATGT	376
QY	361	GGCCGGGGCCAAATGTCTCATTTACCCAGAGTCCCTCCCTACTACCGCTGTGTCTGTAAACAC	420
Db	377	GTCCATGGTGATTGCCTCATTTACCCAGAGCACCCCTACTACCGCTGTGCCTGCAAAATAC	436
QY	421	CCTTACACAGGTCCAGCTGCTCCCAAGTGGTTCCTGTATGCAGGCCAAACCCCTGCCAG	480
Db	437	CCTTACACGGGACCAAGTGTCTCCAAAGTGCTTCGGGATGCGGCCCAACCCCTGCCAGA	496
QY	481	AATGGGGCTACCTGTCTCCCGGATAGCGGAGATCCAGTTCACTGTGCTGTCCCGAC	540
Db	497	ATGGCCGGAGTCTGTTCCCGACACAGACGGAGATCCAGGTTTACCTGTGCCTGTCCAGAC	556
QY	541	CAGTTCAGGGGAAATCTGTGAA-ATAGTTCTGATGACTGCTATGTGGCGATGGCTA	599
Db	557	CAGTATAAGGGGAAATCTGTGAACATAGTCCGGACGACTGTATGTGGGTGATGGCTA	616
QY	600	CTCTTACCGAGGGAAATGAATAGGACAGTCAACAGCATGGCTGCTTTACTGGAATC	659
Db	617	CTCTTACCGAGGGCAAGTGAATAGACAGTTCAACAGAACCCATGCCITTTACTGGAATC	676
QY	660	CCACCTCCTCTTGCAGGAGAAATTACAAC	687
Db	677	CCAGCTCCTCTTGCAGGAGACTTTATAAC	704

Search completed: May 25, 2004, 10:42:53  
Job time : 3061 secs

Sequence 5, Appli  
Sequence 5, Appli  
Sequence 64, Appli  
Sequence 66, Appli  
Sequence 50, Appli  
Sequence 54, Appli  
Sequence 56, Appli  
Sequence 62, Appli  
Patent No. 5200340  
Sequence 10, Appli  
Sequence 38, Appli  
Sequence 42, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 25, Appli  
Patent No. 5200340  
Sequence 3, Appli

28 70.6 4.2 1163 2 US-08-558-269-5  
29 70.6 4.2 1163 3 US-09-410-882-5  
30 70.6 4.2 1170 2 US-08-811-949-64  
31 70.6 4.2 1170 2 US-08-811-949-66  
32 70.6 4.2 1314 2 US-08-811-949-50  
33 70.6 4.2 1314 2 US-08-811-949-54  
34 70.6 4.2 1314 2 US-08-811-949-56  
35 70.6 4.2 1419 2 US-08-811-949-62  
36 70.6 4.2 1738 6 5200340-1  
37 70.6 4.2 1848 3 US-08-814-412-10  
38 70.6 4.2 1974 2 US-08-811-949-38  
39 70.6 4.2 2101 2 US-08-811-949-42  
40 70.6 4.2 2162 1 US-08-113-512-3  
41 70.6 4.2 2162 1 US-08-488-015B-3  
42 70.6 4.2 2162 1 US-08-488-015B-25  
43 67.8 4.0 1724 6 5200340-5  
44 67.8 4.0 2497 6 5185259-2  
45 66.4 3.9 2544 3 US-09-518-046-3

ALIGNMENTS

RESULT 1  
US-08-148-910-14  
; Sequence 14, Application US/08148910  
; Patent No. 5466593  
; GENERAL INFORMATION:  
; APPLICANT: Takeshi SHIMOMURA et al.  
; TITLE OF INVENTION: No. 5466593el Protein and Gene Encoding Said Protein  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch,  
; MEDIUM TYPE: 500 Kb Storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/148,910  
; FILING DATE: No. 5466593ember 5, 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-8850  
; TELEFAX: 202-371-8856  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2033 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: Double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; IMMEDIATE SOURCE:  
; LIBRARY: Pre-made Lambda phage Library,  
; LIBRARY: human liver (49, male) cdna Library (Stratagene)  
US-08-148-910-14

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 05:49:14 ; Search time 91.5 Seconds  
(without alignments)  
10207.457 Million cell updates/sec

Title: US-09-912-559-2

Perfect score: 1683

Sequence: 1 atgtttccaggatgtctga.....aaagtgaagtggcttcttaa 1683

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114.4	6.8	2033	1 US-08-148-910-14	Sequence 14, Appl
2	114.4	6.8	2033	1 US-08-448-937A-14	Sequence 14, Appl
3	93	5.5	970	1 US-08-148-910-3	Sequence 3, Appli
4	93	5.5	970	1 US-08-448-937A-3	Sequence 3, Appli
5	72.2	4.3	329	1 US-08-148-910-2	Sequence 2, Appli
6	72.2	4.3	329	1 US-08-148-910-13	Sequence 13, Appl
7	72.2	4.3	329	1 US-08-448-937A-2	Sequence 2, Appli
8	72.2	4.3	329	1 US-08-448-937A-13	Sequence 13, Appl
9	72.2	4.3	1065	1 US-08-427-640-1	Sequence 1, Appli
10	72.2	4.3	1065	1 US-08-427-640-5	Sequence 5, Appli
11	72.2	4.3	1068	1 US-08-427-640-3	Sequence 3, Appli
12	72.2	4.3	1137	4 US-09-553-498-9	Sequence 9, Appli
13	72.2	4.3	1137	4 US-09-618-868-9	Sequence 9, Appli
14	72.2	4.3	1314	2 US-08-811-949-48	Sequence 48, Appl
15	72.2	4.3	1955	2 US-08-863-795A-39	Sequence 39, Appl
16	72.2	4.3	2457	6 5344773-1	Patent No. 5344773
17	72.2	4.3	2544	4 US-09-703-695A-3	Sequence 3, Appli
18	72.2	4.3	7360	1 US-08-286-740-1	Sequence 1, Appli
19	72.2	4.3	7360	5 PCT-US95-09576-1	Sequence 1, Appli
20	70.6	4.2	1065	2 US-08-811-949-60	Sequence 60, Appl
21	70.6	4.2	1068	1 US-08-137-116-2	Sequence 2, Appli
22	70.6	4.2	1068	1 US-08-427-640-7	Sequence 7, Appli
23	70.6	4.2	1068	2 US-08-811-949-44	Sequence 44, Appl
24	70.6	4.2	1068	2 US-08-811-949-46	Sequence 46, Appl
25	70.6	4.2	1068	2 US-08-811-949-52	Sequence 52, Appl
26	70.6	4.2	1068	2 US-08-811-949-58	Sequence 58, Appl
27	70.6	4.2	1068	6 5223256-3	Patent No. 5223256



Query Match 6.8%; Score 114.4; DB 1; Length 2033;  
Best Local Similarity 49.6%; Pred. No. 3.5e-25;  
Matches 605; Conservative 0; Mismatches 561; Indels 54; Gaps 10;

QY 469 AACCCCTGCAGAAATGGGGTACTGCTCCCGGCATAAGCGGAGATCCAAAGTTCACTGT 528  
DB 742 AGCCCTTGGCTGAACGGGGGCACTGCCACCTGATCGTGGCCACCGGACCAACCGTGTGT 801

QY 529 GCCTGTCCCGACAGTTTCAAGGGGAAATCTGTGAAATAGTTTCTGATGA---CTGCTAT 585  
DB 802 GCCTGCCACACAGCTTTCGTGACGCTCTGCAACATCGAGCTGATGAGCGCTGCTTC 861

QY 586 GTTGGCGATGGCTACTTTACCGAGGGGAAATGAATAGGACATCAACCGAGCAGCGTGC 645  
DB 862 TTGGGGAACGGCACTGGGTACCGTGGCGTGGCGAGCACCTCAGCCTCGGGCTCAGCTGC 921

QY 646 CTTTACTGGAACTCCCACTCTCTTGCAGAGAAATTAACAACATGTTTATGAGGATGCT 705  
DB 922 CTGGCTTGGAACTCCGATCTGCTTACAGAGACTGCAGCTGGCTCGGTGGCGCGCG 981

QY 706 GAAACCCATGGATTTGGGGACACAATTTCTGCAGAAACCCAGATCGGACGAAAGCC 765  
DB 982 GCCCTGTGGCTGGCGCCCATGCTTCTGCCGGAATCCGACAAATGACGAGAGGCC 1041

QY 766 TGGTCTTTTAAAGTTTACCAATGACAGGTGAATGGGAATAGTGTGATGCTCAGCC 825  
DB 1042 TGGTCTTACGT---GGTGAAGGACAGCGGCTCTCTGGGAGTACTCCGCCCTGGAGGCC 1098

QY 826 TGCTCAGCCAGGAGCTTGCCTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTT 885  
DB 1099 TGGCAATCCCT-----CACAGAGTCCAACTGTACCGGATCTCTCTGGCGACCTG 1149

QY 886 CCGGGGTTTGACTCTGTGAAAGACTGAGATAGCAGAGAGGAGATCAAGAGATCTAT 945  
DB 1150 CTTGAGCGAGCTCTCCCGGGCGCGCAGGCTCTCGGAGAGGACAAAGAGAGAGCGTTC 1209

QY 946 GGAGCTTTAAGACAGCGCGGGCAAGCACCCATCGGAGGCGTCCCTCCAGTCTCTCGTG 1005  
DB 1210 CTGGGCGCACATCATCGCGGCTCTCTCTCGCTGCGCGCTCGCACCCCTG----- 1262

QY 1006 CTTGACATCTCCATGCCCGGCGCACTTCTGTGTGGGGCGCTGATCCACCCCTGC 1065  
DB 1263 --GCTGGCGGCATCTACATCGGGGACAGCTTCTGCGCGGAGGAGCCCTGTCCACACCTGC 1320

QY 1066 TGGTGTCTCACTGTGCGCACTGACCGACACATAAAACACAGACATCAA-----AGGTG 1119  
DB 1321 TGGGTGTGTGGCGCGCCCACTGCTCTCCACAGCCCGCCCGGACAGCGTCTCGTG 1380

QY 1120 GTCTAGGAGGACAGGACCTGAAGAAAGAAATTTCTATGAGCAGAGCTTTAGGTCGAG 1179  
DB 1381 GTGTGGGCGACACTTCTTCAACCGCACGACGAGCGTGTGCGCAGACCTTCGGCATCGAG 1440

QY 1180 AAGATATTCAAGTACAGCACTACAAATGAAGAGATGAGATTCGCCCAATGATATGCA 1239  
DB 1441 AAGTACATCCGTACACCTGTACTCGTGTTCACCCCG---CGACACAGCTCTGTC 1497

QY 1240 TTGCTCAAGTTAAAGCCAGTGGATGGTCACTGTGCTTAGAATTCAAATAGTGAAGACT 1299  
DB 1498 CTGATCCGGCTGAAGAAGAGGGAGCGGCTGTGCCACACCTGCGAGTTCTGTCAGGCC 1557

QY 1300 GTGTGCTTGCCTGATG-----GTTCTTTCCTCTGGGAGTGGAGTGCACATCTCTGGC 1353  
DB 1558 ATCTGCTTCCCGAGCCCGGAGCACCTTCCCGGAGGACACAGTGCAGATTCGGGGC 1617

QY 1354 TGGGG-----TGTTACAGAAACAGGAAAGGGTCCCGCCAGCTCTCGATGCCAAAGTC 1407  
DB 1618 TGGGGCCACTTGGATGAGAACGTGAGCGGCTACTCCAGCTCCCTGCGGGAGGCGCTGTC 1677

QY 1408 AAGCTGATTCGCAACACTTTGTGCACTCCCGCACTCTATGACACATGATGATGAC 1467  
DB 1678 CCCCCTGGTCCGCGACCAAGTGTGAGAGCCCTGAGTCTTACGGGCGGACATCAGCCCC 1737

QY 1468 AGTATGATCTGTGAGGAAATCTTCAGAAACCTTGGGCAAGACACCTGCCAGGGTACTCT 1527

DB 1738 AACATGCTCTGTGCGGCTACTT---CGACTGAAGTCCGACGCTGCCAGGGGAGCTCA 1794  
QY 1528 GGAGGCCCTGACTCTGTGAGAGGAGCGGACCTACTTACTGCTCTATGGGATAGTGAAGCTGG 1587  
DB 1795 GGGGGGCCCCCTGGCTCTGCGAGAGAACGGGTGCTTACTCTTACGSCATCATCAGCTGG 1854  
QY 1588 GGCTGGAGTGTGAG-----AAGAGCCAGGGGTCTTACACCAAGTTACAAATTCCTG 1641  
DB 1855 GGTGACGGCTCGGGCGGCTCCACAGCGGGGTCTTACACCGCGGTGGCGCAACTATGTG 1914

QY 1642 AATTGGATCAAGCCACCAT 1661  
DB 1915 GACTGGATCAACGACCGGAT 1934

RESULT 2  
US-08-448-937A-14  
; Sequence 14, Application US/08448937A  
; Patent No. 5677154  
; GENERAL INFORMATION:  
; APPLICANT: Takeshi SHIMOMURA et al.  
; TITLE OF INVENTION: No. 5677164e1 Protein and Gene Encoding Said Protein  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch,  
; MEDIUM TYPE: 500 Kb Storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/448,937A  
; FILING DATE: May 24, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/148,910  
; FILING DATE: No. 5677164ember 5, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-8850  
; TELEFAX: 202-371-8856  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2033 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; IMMEDIATE SOURCE:  
; LIBRARY: Pre-made Lambda phage Library,  
; LIBRARY: human liver(49, male) cDNA Library (Stratagene)  
; US-08-448-937A-14

Query Match 6.8%; Score 114.4; DB 1; Length 2033;  
Best Local Similarity 49.6%; Pred. No. 3.5e-25;  
Matches 605; Conservative 0; Mismatches 561; Indels 54; Gaps 10;

QY 469 AACCCCTGCAGAAATGGGGTACTGCTCCCGGCATAAGCGGAGATCCAAAGTTCACTGT 528  
DB 742 AGCCCTTGGCTGAACGGGGGCACTGCCACCTGATCGTGGCCACCGGACCAACCGTGTGT 801

QY 529 GCCTGTCCGACAGTTCCAGGGGAAATTCGTGAAATAGTTCCTGATGA---CTGCTAT 585  
Db 802 CCCTGCCCAACAGAGCTTCGCTGGAGCGCTCTGCAACATCGAGCTGATGAGCGCTGTTTC 861  
QY 586 GTTGGCGATGCTACTCTTACCGAGGGGAAATGAATAGGACAGTCAACCGACGATGCGTGC 645  
Db 862 TTGGGGAACGGCAGCTGGGTACCGTGGGTGGTGGCCAGCACCTCAGCCTCGGGCCTCAGCTGC 921  
QY 646 CTTTACTGGAATCCCACTCCTCTTTCAGAGAGAAATACAACTGTTTATGGAGATGCT 705  
Db 922 CTGGCTGGAACTCGATCTGCTTACAGAGAGCTGACGTGGACTCGTGGGCGCGCG 981  
QY 706 GAAACCCATGGATTTGGGAAACAAATTTCTGAGAAACCCAGATGCGGACGAAAGGCC 765  
Db 982 GCCCTGTGGGCTGGGCGCCCATGCTTACTGCGGAATCGGGAATGACGAGAGGCC 1041  
QY 766 TGGTGTCTTATTAAGTTACCAATACAAAGTGAATGGGAATCTGATGTTCTCAGCC 825  
Db 1042 TGGTGTCTAGT---GGTGAAGAGACAGCGGCTCTCTGGGAGTACTCGCGCTGGAGGCC 1098  
QY 826 TGCTCAGCCAGGAGTTGGCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTT 885  
Db 1099 TGGGAATCCCT-----CACAGAGTCCAACTGTCAACGGATCTCTGGGACCGCTG 1149  
QY 886 CCGGGTTGACTCTGTGGAAGACTGAGATAGCAGAGAGAGATCAGAGAACTAT 945  
Db 1150 CTTGAGCAGGCTCCCGGGCGCGCAGGCTGCGGAGGAGGCAAGAGAGAGCTTC 1209  
QY 946 GGAGGCTTTAAGACACGGCGGGCAAGCACCCATGCGAGGGCTCCCTCCAGTCTCTCGGTG 1005  
Db 1210 CTGGGCGACGATATCATCGGGCTCTCTCTCGTGGCGGCTCGACCCCTG----- 1262  
QY 1006 CTTCTGACATCTCATGCCCCAGGGCCACTTCTGTGTGGGGGCTGATCCACCCCTGC 1065  
Db 1263 --GCTGGCGCCATCTATCATCGGGGACAGCTTCTCGCGGAGGCTGTGTCACACCTGC 1320  
QY 1066 TGGGTGCTCACTGCTGCCACTGCAACGACATATAAAGACAGACATCTAA-----AGGTG 1119  
Db 1321 TGGGTGCTGCGCGCCACCTGCTCTCTCCACACCCCGGAGGAGCGTCTCCGTG 1380  
QY 1120 GTGTAGGGGACAGGACCTGAAGAAAGAAATTTCAATGACAGAGCTTTAGGGTGAG 1179  
Db 1381 GTGTGGGCGACACTTCTTCAACCGCACGAGGAGCTGACGACACTTCGGCATCAG 1440  
QY 1180 AAGATATTCAAGTACAGCCACTACATGAAGAGATGAGATTCGCCCAATGATATTGCA 1239  
Db 1441 AAGTACATCCGTTACACCTGTACTCGGTGTTCAACCCAG---CGACACGACCTCTC 1497  
QY 1240 TTGCTCAAGTTAAGCCAGTGGATGCTACTGTCTTAGAATCCAAATAGTGAAGACT 1299  
Db 1498 CTGATCCGCTGAAGAAAGAGGACCGCTGTGCCACACGCTCGAGTTCTGTGAGGCC 1557  
QY 1300 GTGTGCTTGCCTGATG-----GGTCTTCTCTGGAGTGAGTGCCACATCTCTGCG 1353  
Db 1558 ATCTGCTCCGAGCCCGGAGCACCTTCCCGGAGGACACAGTGGCCAGATTGCGGGC 1617  
QY 1354 TGGGG-----TGTTACAGAAACAGGAAAGGTCCTCGGACGCTCTGATGCCAAAGTC 1407  
Db 1618 TGGGGCCACTTGGATGAGAACGTTGAGCGGCTACTCCAGCTCCTCGGGAGGCCCTGTGTC 1677  
QY 1408 AAGCTGATGCCAACHTTGTGCAACTCCGCCAACTCTATGACCAATGATGATGAC 1467  
Db 1678 CCGCTGGTCCGACCAAGTGCAGACGCTTGAAGTCTACGGGCGGACATCAGCCCC 1737  
QY 1468 AGTATGATCTGTGAGGAATCTTCAGAAACCTGGGCAAGACACCTGCCAGGGTGACTCT 1527  
Db 1738 AACATGCTCTGTGCGGCTACTT---CGACTGCAAGTCCGAGCGCTGCCAGGGGACTCA 1794  
QY 1528 GAGAGGCCCTGACTGTGAGAGAGGACGGACCTACTAGTCTATGGGATAGTGAAGCTGG 1587  
Db 1795 GGGGGGCCCCGTGCGGCTCGAGAGAAAGCGGCTTACCTTCTACGGCATCATCAGCTGG 1854

QY 1588 GGCCTGGAGTGTGAG-----AAGAGGCCAGGGGCTTACACCCAAAGTTACCAATTCCTG 1641  
Db 1855 GGTGACGGCTGGGGCGGCTCCACAAGCGGGGGTCTTACACCCGCGTGGCCAACTATGTG 1914  
QY 1642 AATTGGATCAAGGCCACCAT 1661  
Db 1915 GACTGATCAACGACCGGAT 1934  
RESULT 3  
US-08-148-910-3  
; Sequence 3, Application US/08148910  
; Patent No. 546593  
; GENERAL INFORMATION:  
; APPLICANT: Takeshi SHIMOMURA et al.  
; TITLE OF INVENTION: No. 546593el Protein and Gene Encoding Said Protein  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch,  
; MEDIUM TYPE: 500 Kb Storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WordPerfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/148,910  
; FILING DATE: No. 546593ember 5, 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-8850  
; TELEFAX: 202-371-8856  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 970 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; IMMEDIATE SOURCE:  
; LIBRARY: Pre-made Lambda phage library, human liver (49, male)  
; LIBRARY: cDNA Library (Stratagene)  
; US-08-148-910-3

Query Match 5.5%; Score 93; DB 1; Length 970;  
Best Local Similarity 52.7%; Pred. No. 1.1e-18;  
Matches 357; Conservative 0; Mismatches 290; Indels 30; Gaps 6;

QY 1009 CTGACATCTCCATGCCCCAGGCCACTTCTGTGGGGCGCTGATCCACCCCTGCTGG 1068  
Db 201 CTGGCCGCCATCTACATCGGGGACAGCTTCTGCGCGGGAGCCTGGTCCACACTGCTGG 260  
QY 1069 GTGCTCACTGTGCGCCACTGCACCGACATATAAACACAGACA-----TCTAAAGGTGGTG 1122  
Db 261 GTGGTGTGCGCGCGCCACTGCTTCTCCACAGCCCCCAGGAGACAGCTCTCCGTGGTG 320  
QY 1123 CTAGGGACACGAGACCTGAAGAAAGAAATTTTCATGAGCAGAGCTTAGGGTGCAGAG 1182

Db 321 CTGGGCGCAGCATTCTTCAACCGCAGCGGACGGACGTGACGCAGACCTTCGGCATCGAGAAG 380  
 QY 1183 ATATTCAGTACAGCCACTTACAAATGAAGAAGATGAGATTCGCCAATGATATGCAATTG 1242  
 Db 381 TACATCCCTGACACCTGTACTCGGTGTTCACCCCA---CGGACCGACGACCTGCTCTG 437  
 QY 1243 CTCAGATTAAAGCCAGTGTGCTACTGCTCTAGATCCCAATACGTCGACGATGTG 1302  
 Db 438 ATCCGGCTCAAGAAAGAGGGACCGCTGTGCCACACGCTCGAGTTCGTGACCCATC 497  
 QY 1303 TGCCTTGCCTGAG---GTCCTTTCCCTCTGGAGTGTGCTGACATCTCTGCTGG 1356  
 Db 498 TGCCTTGCCTGAGCCCGGACGACCTTCCCGCAGGACACAGTGCAGATTCGGGGCTGG 557  
 QY 1357 GG-----TGTTACAGAAACAGGAAAGGTCCTCCGCGCAGCTCTCTGGATGCCAAGTCAAG 1410  
 Db 558 GGCACCTTGGATGAGAAAGTGTGAGCGGTACTCTCGAGCTCCCTGCGGAGGCCCTGCTCCC 617  
 QY 1411 CTGATTGCCAAGCATTGTGTGACCTCCGCCCACTCTATGACCAATGATGATGACAGT 1470  
 Db 618 CTGGTGGCGGACCAAGTGTGAGCAGCCCTGAGGTCTACGGCGCGGCATCAGCCCCAAC 677  
 QY 1471 ATGATCTGTGCAAGAAATCTTTCAGAAACCTGGGCAAGACCTGCCAGGGTGAATCTGGA 1530  
 Db 678 ATGCTCTGTGCGGCTACTT---CGACTGCAAGTCCGAGCGCTGCCAGGGGACTCAGGG 734  
 QY 1531 GCGCCCTGACTGTGAGAGGACGGGACCTACTAGTCTATGGGATGATGAGCTGGGGC 1590  
 Db 735 GGGCCCTTGGCTGTGAGAGAGAGCGGTGTACCTCTACGGGCATCATCAGCTGGGGT 794  
 QY 1591 CTGGAGTGTGAG-----AAGAGGCGGAGGGTCTACACCCCAAGTTACCAATTCCTGAAT 1644  
 Db 795 GACGGCTGGGCGGCTCCACAAAGCGGGGTCTACACCGCGTGGCCACTATGTGGAC 854  
 QY 1645 TGGATCAAGCCACCAT 1661  
 Db 855 TGGATCAAGCCGGAT 871

RESULT 4

US-08-448-937A-3

; Sequence 3, Application US/08448937A

; Patent No. 567164

; GENERAL INFORMATION:

; APPLICANT: Takeshi SHIMOMURA et al.

; TITLE OF INVENTION: No. 567164el Protein and Gene Encoding Said Protein

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSES: Wendeth, Lind & Penack

; STREET: 805 Fifteenth Street, N.W., #700

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch,

; MEDIUM TYPE: 500 Kb Storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/448,937A

; FILING DATE: May 24, 1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/148,910

; FILING DATE: No. 567164member 5, 1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Warren M. Cheek, Jr.

; REGISTRATION NUMBER: 33,367

; REFERENCE/DOCKET NUMBER:

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-371-8850

; TELEFAX: 202-371-8856  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 970 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: human  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: Pre-made Lambda phage library, human liver (49, male)  
 ; LIBRARY: cDNA Library (Stratagene)  
 ; US-08-448-937A-3

Query Match 5.5%; Score 93; DB 1; Length 970;

Best Local Similarity 52.7%; Pred. No. 1.1e-18;

Matches 357; Conservative 0; Mismatches 290; Indels 30; Gaps 6;

QY 1009 CTGACCATCTCTCATGTCCTCCAGGGCCACTTCTGTGTGGGCGCTGATCCACCCCTCTG 1068  
 Db 201 CTGGCGCCATCTACATCGGGGACAGCTTCTGCGCGGGAGCTGTCCACACCTCTG 260  
 QY 1069 GTGTCACTGTGCTCCACTGACCGACGATATAAAACACAGACA-----TCTAAAGTGGT 1122  
 Db 261 GTGTGTGGCGCGCCACTCTCTTCCACAGCCCCCAGGACAGCGTCTCCGTTGGT 320  
 QY 1123 CTAGGGACACGAGGACCTTGAAGAAAGAAATTTTCATGACGACAGCTTTAGGGTGCAGAA 1182  
 Db 321 CTGGCCAGCAGCTTCTTCAACCGCAGCGAGCGTGCAGCAGACCTTCGCGATCGAGA 380  
 QY 1183 ATATTCAGTACAGCCACTTACAATGAAGAGATGAGATTTCCCAATGATATGCAATTG 1242  
 Db 381 TACATCCCGTACACCTCTGTACTCGGTGTTCAACCCCA---GGACACACCTCTGCTCT 437  
 QY 1243 CTCAAGTTAAAGCCAGTGGTGTGCTCTAGATCTTCAAAATACGTAAGAGTGTG 1302  
 Db 438 ATCCGGCTGGAAGAAAGGAGCGGACCGCTGTGCCACACGCTTCGTCAGGCCATC 497  
 QY 1303 TGCTTGGCTGATG-----GGTCTTCTCTCTGGAGTGTGCGCACATCTCTGCTGG 1356  
 Db 498 TGCCTGCGGAGCGCGGACGACCTTCCCGCAGGACACAAGTCCAGATTCGCGGCTGG 557  
 QY 1357 GG-----TGTTACAGAAACAGGAAAGGGTCCCGCAGCTCTCTGGATGCCAAGTCAAG 1410  
 Db 558 GGCCACTTGGATGAGAACGTGAGCGGTACTCTCCAGCTCTCCGCGAGGCCCTGTTCC 617  
 QY 1411 CTGATTGCCACACTTTGTGCAACTCCGCCACTCTATGACCAATGATGATGACAGT 1470  
 Db 618 CTGTGCGCCACCAAGTGTGAGAGCGCTGTAGGTCTACGGCGCCGACATCAGCCCCAAC 677  
 QY 1471 ATGATCTGTGAGGAAATCTTTCAGAAACCTGGGCAAGACACCTGCCAGGGTGAATCTGGA 1530  
 Db 678 ATGCTCTGTGCGGCTACTT---CGACTGCAAGTCCGAGCGCTGCCAGGGGACTCAGGG 734  
 QY 1531 GGCCCTGACCTGTGAGAGGAGCGGACCTACTACTGCTATGGGATGATGAGCTGGGGC 1590  
 Db 735 GGGCCCTTGGCTGTGAGAGAGAGCGGCTGTACTTCTACGGCATCATCAGCTGGGT 794  
 QY 1591 CTGGAGTGTGAG-----AAGAGGCGGAGGGTCTACACCCCAAGTTACCAATTCCTGAAT 1644  
 Db 795 GACGGCTGGGCGGCTCCCAAGCGGGGTCTACACCGCGTGGCCAACTATGTGGAC 854  
 QY 1645 TGGATCAAGCCACCAT 1661  
 Db 855 TGGATCAAGCCGGAT 871

RESULT 5

US-08-148-910-2

; Sequence 2, Application US/08148910

; Patent No. 5466593

Query Match	4.3%;	Score 72.2;	DB 1;	Length 329;
Best Local Similarity	59.3%;	Fred. No. 1.7e-12;		
Matches 163;	Conservative	0;	Mismatches 103;	Indels 9; Gaps 2;
QY	1381	TCCGCGCAGCTCCTGGATGCGAAAGTCAAGCTGATTGCCAACACTTTGTGTCAACTCCCGC	1440	
Db	52	TCCAGTCCCTCGGGGAGGCCCTGGTCCCTGGTCCGCGACCAACAAGTGCAGCAGCCCT	111	
QY	1441	CAACTCTATGACCACATGATTGATGACAGTATCATCTGTGCAGGAATCTTCAGAAACCT	1500	
Db	112	GAGGTCTACGGCGCGACATCAGCCCCACATGCTCTGTGCGGCTACTT---CGACTGC	168	
QY	1501	GGGCAAGACACCTGCCAGGCTGACTCTGGAGGCCCTGACTGTGTGAAGAGACGGCACC	1560	
Db	169	AAGTCCGACCCCTGCCAGGGGGACTCAGGGGGGCCCTGGCCCTCGAGAGAAAGCGCGTG	228	
QY	1561	TACTAGCTCTATGGGTAGTGAGCTGGGGCTTGAGTGTGAG-----AAGAGGCCAGGG	1614	
Db	229	GCTTACTCTACGGCATCATAGCTGGGTGACGGCTCGGGCGGCTCCACAGCCGGG	288	
QY	1615	GCTACACCCAAAGTATACCAATTCCCTGAATTGGAT	1649	
Db	289	GTTCAACCCGCGCTGGCCAACTATGTGGACTGGAT	323	

RESULT 7  
US-08-448

	Query Match	4.3%;	Score 72.2;	DB 1;	Length 329;
	Best Local Similarity	59.3%;	Pred. No. 1.7e-12;		
	Matches 163;	Conservative	0;	Mismatches 103;	Indels 9; Gaps 2;
QY	1381	TCGCGCCAGCTCCTGGATGCCAAAGTCAAGCTGATTGCCAACACTTTGTGTGCAACTCCCGC	1440		
DB	52	TCCAGCTCCCTGCGGAGGCCCTGGTGCCTCGTGCCTGCGCGACCAAGAGTGCAGACCCCT	111		
QY	1441	CAACTCTATGACCACATGATTGATGACAGTATGATCTGTGCAGGAAATCTTCAGAAACCT	1500		
DB	112	GAGGTCCTAGCGGCCGACATCAGCCCCAACATGCTGTGCGGCTACTT---	CGACTGC	158	
QY	1501	GGGCAAGACACTGCCAGGGTGACTCTGGAGGCCCCCTGACTGTGAGAGAGCGGCACC	1560		
DB	169	AAGTCCGACGCCCTGCCAGGGGACTCAGGGGGGCCCTGGCTTCGAGAGAAGACGCCGCTG	228		
QY	1561	TACTAGCTCTATGGGATAGTGAAGCTGGGCGCTGGAGTGTGAG-----	AAGAGGCCCAGGG	1514	
DB	229	GCTTACTCTCTACGGCATCATCAGCTGGGTGACGGCTGGGGCGGCTCCACAGCCGGG	288		
QY	1615	GTCTACACCCCAAGTTACCAAAATTCCTGAATTGGAT	1649		
DB	289	GTCTACACCCGCGTGGCCAACTATGTGGAGTGGAT	323		

Sequence 2, Application US/08448937A  
Patent No. 5677164  
GENERAL INFORMATION:  
APPLICANT: Takeshi SHIMOMURA et al.  
TITLE OF INVENTION: No. 5677164el Protein and Gene Encoding Said Protein  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch,  
MEDIUM TYPE: 500 Kb Storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/448,937A  
FILING DATE: May 24, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/148,910  
FILING DATE: No. 5677164ember 5, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX: 202-371-8856  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 329 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORGANISM: human  
IMMEDIATE SOURCE:  
LIBRARY: Quick-cloneTM human liver cDNA (Clontech)  
US-08-448-937A-2

Query Match 4.3%; Score 72.2; DB 1; Length 329;  
Best Local Similarity 59.3%; Pred. No. 1.7e-12;  
Matches 163; Conservative 0; Mismatches 103; Indels 9; Gaps 2;  
QY 1381 TCCGCGAGCTCCTGGATGCCAAAGTCAAGCTGATTGCCAACACTTTGTGCAACTCCCGC 1440  
Db 52 TCCAGCTCCCTGCGGGAGGCCCTGCTCCCTGCTGCGCGACCAAGTCAGCAGCCCT 111  
QY 1441 CAACCTATGACCATGATTGATGACAGTATGATCTGTGAGGAAATCTTCAGAAACCT 1500  
Db 112 GAGGTCTACGGCGCGGACATCAGCCCCAACATGCTCTGTGCGGCTACTT---CGACTGC 168  
QY 1501 GGGCAAGACACTGCGAGGCTGACTCTGAGGCCCCCTGACTGTGAGAGAGCGGCACC 1560  
Db 169 AAGTCCGAGCGCTGCCAGGGGACTCAGGGGGCCCCCTGGCTCGGAGAGACGGCGTG 228  
QY 1561 TACTACGCTTATGGGATAGTGAAGTGGGCTGTGAGTGTGAG-----AAGAGGCCAGGG 1614  
Db 229 GCTTACCTCTACGGCATCATCAGTGGGGTACGCGCTGCGGGCGGCTCCCAAGCCGGG 288  
QY 1615 GTCTACACCCAGTTACCAAAATTCCTGAATTGGAT 1649  
Db 289 GTCTACACCCGGTGGCCAACTATGTGGACTGGAT 323

RESULT 8

US-08-448-937A-13  
Sequence 13, Application US/08448937A  
Patent No. 5677164  
GENERAL INFORMATION:  
APPLICANT: Takeshi SHIMOMURA et al.  
TITLE OF INVENTION: No. 5677164el Protein and Gene Encoding Said Protein  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch,  
MEDIUM TYPE: 500 Kb Storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/448,937A  
FILING DATE: May 24, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/148,910  
FILING DATE: No. 5677164ember 5, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX: 202-371-8856  
TELEX:  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 329 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORGANISM: human  
IMMEDIATE SOURCE:  
LIBRARY: Quick-cloneTM human liver cDNA (Clontech)  
US-08-448-937A-13

Query Match 4.3%; Score 72.2; DB 1; Length 329;  
Best Local Similarity 59.3%; Pred. No. 1.7e-12;  
Matches 163; Conservative 0; Mismatches 103; Indels 9; Gaps 2;  
QY 1381 TCCGCGAGCTCCTGGATGCCAAAGTCAAGCTGATTGCCAACACTTTGTGCAACTCCCGC 1440  
Db 52 TCCAGCTCCCTGCGGGAGGCCCTGCTCCCTGCTGCGCGACCAAGTCAGCAGCCCT 111  
QY 1441 CAACCTATGACCATGATTGATGACAGTATGATCTGTGAGGAAATCTTCAGAAACCT 1500  
Db 112 GAGGTCTACGGCGCGGACATCAGCCCCAACATGCTCTGTGCGGCTACTT---CGACTGC 168  
QY 1501 GGGCAAGACACTGCGAGGCTGACTCTGAGGCCCCCTGACTGTGAGAGAGCGGCACC 1560  
Db 169 AAGTCCGAGCGCTGCCAGGGGACTCAGGGGGCCCCCTGGCTCGGAGAGACGGCGTG 228  
QY 1561 TACTACGCTTATGGGATAGTGAAGTGGGCTGTGAGTGTGAG-----AAGAGGCCAGGG 1614  
Db 229 GCTTACCTCTACGGCATCATCAGTGGGGTACGCGCTGCGGGCGGCTCCCAAGCCGGG 288  
QY 1615 GTCTACACCCAGTTACCAAAATTCCTGAATTGGAT 1649  
Db 289 GTCTACACCCGGTGGCCAACTATGTGGACTGGAT 323

RESULT 9  
US-08-427-640-1  
; Sequence 1, Application US/08427640  
; Patent No. 5658788  
; GENERAL INFORMATION:  
; APPLICANT: Berg et al.  
; TITLE OF INVENTION: Compounds and Methods for Treatment of Thromboembolic Disorder  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: IN.  
; COUNTRY: U.S.A.  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: Macintosh  
; SOFTWARE: Microsoft Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/427,640  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/689,410  
; FILING DATE: 22 APRIL 1991  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1065 base pairs  
; TYPE: Nucleic acid  
; STRANDEDNESS: Double  
; TOPOLOGY: Linear  
; MOLECULE TYPE: DNA  
US-08-427-640-1

Query Match 4.3%; Score 72.2; DB 1; Length 1065;  
Best Local Similarity 50.5%; Pred. No. 3.7e-12;  
Matches 340; Conservative 0; Mismatches 288; Indels 45; Gaps 5;

QY	1028	AGAGGCCACTTCTGTGGTGGGGCGCTGATCCACCCCTGCTGGGTGCTCAGTCTGCCACT	1087
DB	392	AGCGGTTCTGTGCGGGGGGATCTATCAGTCTCTGTGGATCTCTTGGCGCCACT	451
QY	1088	GCACCGA-----CATAAAAACCAAGATCTAAAGTGGTCTAGGGGACCAAGCACTGA	1141
DB	452	GCATTCCAGGAGAGGTTTTCCGCCCCACACCTGACGGTGATCTTGGGCAGAACATACCGGG	511
QY	1142	AGAAGAAGAAATTCATGAGCAGAGCTTTAGGGTGCGAAGATATTCAGGTACAGCCACT	1201
DB	512	TGCTTCCTGCGGAGGAGGAGGAGAAATTTGAAGTCGAAAAATACATTTGTCATAAGGAAT	571
QY	1202	ACAATGAAGAGATGAGATTTCCCCACCAATGATATTGCAATTGCTCAAAGTTAAAGCCAGTGG	1261
DB	572	TCGAT-----GATGACACITACGACATGACATTTGCGCTGCTGCAGCTGAAATCGGATT	625
QY	1262	ATGGTCACTGTGCTTAGAATCCAAATACGTGAAGACATGTGTGCTTGCCTGATGGGTCCT	1321
DB	626	CGTCCCGCTGTGCCCAGGAGAGACGCTGTGTCGCACTGTGTGCTTTCCCGCGCGGACC	685
QY	1322	TTCCCT-----CTGGGAGTCAGTCCCACTCTCTGGCTGGGGTGTACAGAA	1369
DB	686	TGAGCTGCCGAGCTGACGGAGTGTGAGCTCTCCGGCTACGGCAAGCATGAGGCTTGT	745
QY	1370	CAGGAAAGGGTCCCGCCAGCTCCTGGATGCAAAAGTCAAGCTGATTGCCAACACTTTGT	1429
DB	746	CTCCTTCTATTTCGGAGCGGCTGAAGGAGGCTCATGTCCAGACTGTACCATCCAGCGCT	805
QY	1430	GCACTCCCGCACTCTATGACCACATGATTGATGACAGTATGATCTGTGCAGGAATC	1489
DB	806	GCACTACAAACATTTACTTTAACAGAACAGTCCACGACCAACATGCTGTGTGCTCGAGACA	865
QY	1490	TTCAGAAACCTGG-----GCAAGACACCTGCCAGGGTGACTCTCGAGGCC	1534





Query Match	4.3%;	Score 72.2;	DB 4;	Length 1137;
Best Local Similarity	50.5%;	Pred. No. 3.8e-12;		
Matches 340;	Conservative 0;	Mismatches 288;	Indels 45;	Gaps 5;
1028	AGGGCCATTCTGTGGTGGGGCGCTGATCCACCCTCTGTGGGTGCTCACTGTCGCCACT	1087		
461	AGCGGTCTCTGTGCGGGGCACTCATCAGTCTCTCTGTGGATTCTCTTCGCCGCCACT	520		
1088	GCACCGA-----CATAAAAACAGACATCTAAAGTGTGTAGGGGACCAAGCACTGA	1141		
521	GTTTCAGGAGAGGTTTCGCGCCCAACCCTGACGGTGATCTTGGCGAGAACATCCGG	580		
1142	AGAAAGAGAATTCATGAGCGAGGCTTTAGGGTCAGAGAATATTCAGTAGACGCCT	1201		
581	TGTCCTCTGGCGAGGAGCAGAAATTTGAAGTCGAAAAATACATTTGCCATAAGGAA	640		
1202	ACAAATGAAGAGATGAGATTCGCCCAATGATATTGCATTTGCTCAAGTTTAAAGCAGTGG	1261		
641	TCGAT-----GATGACATTACGACATGACATTCGCTGTGCMGCTGAAATCGGATT	694		
1262	ATGGTCACTGTGCTCTAGAAATCCAAATACGTGAAGACTGTGCTTTGCCGTGATGGGTCT	1321		
695	CGTCCCGCTGTGCCAGGAGACGCGTGTGCGCACTGTGTGCTTCCCGCGGAC	754		
1322	TTCCCT-----CTGGGAGTAGTGCCACATCTCTGCTGGCTGGGTGTACAGAAA	1369		
755	TGCAGCTGCCGACTGGAACGAGTGTGAGCTCTCCGGCTACGSCAAGCATGAGGCTTTGT	814		
1370	CAGGAAAAGGGTCCCGCAGCTCTCGATGCGCAAGCTCAAGCTGATTGCCAACACTTTGT	1429		
815	CTCCTTTTATTTCGAGCGGCTGAAGAGGCTCATGTACAGCTTACCCATCCAGCGCT	874		
1430	GCAACTCCCGCCAACTCTATGACACATGATTGATGACAGTATGATCTGTGCAGGAAATC	1489		
875	GCACATCAACAATTTACTTTAACAGAACAGTCCCGACACATGCTGTGTCTGGAGACA	934		
1490	TTCAGAAAACCTGG-----GCAAGACACTTCGAGGTGACTCTGGAGGCC	1534		
935	CTCGAGCGCGCGGCCCCAGGCAAACTTGACAGCCCTGCCAGGGCGATTCCGGAGGCC	994		
1535	CCCTGACCTGTGAGAAGGACGGCACTTACTACTCTATGGGATAGTGAGCTGGGGCTGG	1594		
995	CCCTGTGTCTTGAACGATGGCCGATGACTTTGGTGGGCATCATCAGCTGGGGCTGG	1054		
1595	AGTGTGAGAGAGG-----CCAGGGGTCTACACCCCAAGTTACAAATTCCTGAATTGGA	1648		
1055	GCTGTGCAGAGAGGATGTCGCGGTGTGTACACCAAGGTTACCAACTACCTAGACTGGA	1114		
1649	TCAAAGCCACCAT	1661		
1115	TTCGTGACAAACAT	1127		

```

; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1137)
US-09-618-869-9

Query Match          4.3%; Score 72.2; DB 4; Length 1137;
Best Local Similarity 50.5%; Pred. No. 3.8e-12;
Matches 340; Conservative 0; Mismatches 288; Indels 45; Gaps 5;

QY 1028 AGGCGCACTTCTGTGTGGTGGGCGCTGATCCACCTCTGCTGGGTGCTCACTGCTGCCACT 1087
    |||
Db 461 AGCGGTCTCTGTGTGGGCGCATACACTCAGCTCTCTCTGTGGATTTCTCTGCGCCACT 520
    |||

QY 1088 GCACCGA-----CATATAAAACAGACATCTAAAGTGTGTAGGGGACCAAGCACTGA 1141
    |||
Db 521 GCTTCCAGGAGAGGTTTCCGCCCCACCACCTCAGCGTGATCTTGGGAGAAACATACCGG 580

QY 1142 AGAAGAGAAGATTTCATGACGAGAGCTTTAGGTCGAGAGATATTCAGTACAGCCACT 1201
    |||
Db 581 TGTCTCCTCGGAGGAGGAGAGAGAAATTTGAAGTGGAAAAATACATGTCTCATAGGAAT 640

QY 1202 ACAATGAAAGATGAGATTCGCCACAATGATATTCATTTGCTCAAGTTAAAGCCAGTGG 1261
    |||
Db 641 TCGAT-----GATGACACTTACGACAATGACATTGCGCTGTGCACTGAAATCGAATT 694

QY 1262 ATGTCACCTGTGCTTAGAATCCAAATACGTGAAGACTGTGCTTGCTGTGCTGATGGGTCC 1321
    |||
Db 695 CPTCCCGCTGTGCCAGGAGAGCGCGTGTTCGCACTGTGTGCTTCCCCGCGCGACC 754

QY 1322 TTCCTC-----CTGGGAGTGAAGTGCACATCTCTGGCTGGGTGTTTACAGAA 1369
    |||
Db 755 TGCAGCTGCCGACTCGACGGAGTGTGAGCTCTCCGGCTACGGCAAGCATGAGCGCTTGT 814

QY 1370 CAGGAAAGGGTCCCGCAGCTCTTGATGCCAAAGTCAAGCTGATTGCCAACACTTTGT 1429
    |||
Db 815 CTCCTTTCTATTTCGAGCGGCTGAAGAGGCTCATGTCAAGCTGTACCCATCAGCGCT 874

QY 1430 GCAACTCCGCGCAACTCTATGACCAATGATGACAGTATGATCTGTGCAAGGAAATC 1489
    |||
Db 875 GCACATCACAACTTTACTTAAACAGAACATGACCCGACAACTGCTGTGTGCTGGAGACA 934

QY 1490 TTCAGAAACCTGG-----GCAACACACCTGCCAGGGTGAATCTCTGGAGGCC 1534
    |||
Db 935 CTGGAGCGCGCGGCCAGGCAAACTTGACACACGCTGCCAGGGCGATTCGGAGGCC 994

QY 1535 CCCTGACCTGTGAGAGGAGCGCACCTTACTACCTCTATGGGATAGTGAAGTGGGCGCTGG 1594
    |||
Db 995 CCCTGTGTGTCTGAACAGATGGCGCGATGACTTTTGGTGGGCATCATCAGCTGGGCGCTGG 1054

QY 1595 AGTGTGAGAGAGG-----CCAGGGGTCTACACCCAAAGTTACCAATTCCTGAATTGGA 1648
    |||
Db 1055 GCTGTGACAGAGGATGTCCTCGGGTGTGTACACCAAGGTTACCAACTACTAGACTGGA 1114

QY 1649 TCAAGCCACCAT 1661
    |||
Db 1115 TTGCTGACAAACAT 1127
    |||

```

NUMBER OF SEQUENCES: 67  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUBSTADT,  
 ADDRESSEE: P.C.  
 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
 CITY: ARLINGTON  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22202

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 FILING DATE: 05-MAR-1997  
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
 NAME: OBLON, NORMAN F.  
 REGISTRATION NUMBER: 24,618  
 REFERENCE/DOCKET NUMBER: 19-966-0  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 703-413-3000  
 TELEFAX: 703-413-2220

INFORMATION FOR SEQ ID NO: 48:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1314 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: circular  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1311

US-08-811-949-48

Query Match 4.3%; Score 72.2; DB 2; Length 1314;  
 Best Local Similarity 50.5%; Pred. No. 4.2e-12;  
 Matches 340; Conservative 0; Mismatches 288; Indels 45; Gaps 5;

Qy	1028	AGGCCATTCTGTGGTGGGGGCTGATCCACCCCTGCTGGTCTCTCACTGCTCCCACT	1087
Db	638	AGCGGTTCTGTGGGGGGCTACTATCAGCTCTCTGATGATCTCTGCGCCCACT	697
Qy	1088	GCACCGA-----CATAAAAACAGACATCTAAAGGTGGTCTAGGGACACGAGCTGA	1141
Db	698	GCTTCCAGGAGAGGTTTCCGGCCCAACACCTGACCGTGATCTTGGCAGAACATACCGGG	757
Qy	1142	AGAAAGAGATTTTCATGACGAGAGCTTTAGGGTGCAGAGATATTCAAGTACAGCCACT	1201
Db	758	TGCTCCCTGGGAGGAGGAGAGAGAAATTTGAAGTCGAAAAATACATTTCTCATTAGGAT	817
Qy	1202	ACAATGAAGAGATGAGATTCCCAATGATATTGCTCAAGTTAAAGCCAGTGG	1261
Db	818	TCGAT-----GATGACATTAGCAATGACATTCGCTGCTGAGCTGAAATCGGATT	871
Qy	1262	ATGGTCACTGTGCTTAGAATCAATAGTGAAGACTGTGCTTGTGCTGATGGTCT	1321
Db	872	CGTCCCGCTGTGCCAGGAGAGACGGTGGTCCGCACTGTGTGCTTCCCGCGGGACC	931
Qy	1322	TTCCCT-----CTGGAGTGAAGTGCACATCTCTGGCTGGGGTGTACAGAAA	1369
Db	932	TGCAGTGGCGGAGCTGACGGAGTGTGAGCTCTCCGGCTACGGCAAGCATGAGGCTTGT	991
Qy	1370	CAGGAAAAGGTCCCGCCAGCTCTGGATGCGCAAGTCAAGTGAATTCGCAACATTTGT	1429
Db	992	CTCCTTTCTATTGGGAGCGGCTGAAGAGGCTCATGTGACAGTGTACCCATCCAGCCGCT	1051
Qy	1430	GCAACTCCCGCACTCTATGACCATGATTGATGACAGATGATCTGTGTCAGGAAATC	1489
Db	1052	GCATATCAACAACTTACTTAAACAGACAGTACCAGACACATGCTGTGTGTCGAGACA	1111

Qy	1490	TTCAGAAACCTGG-----GCAAGACACCTGCCAGGTGACTCTTGAGGCC	1534
Db	1112	CTGGAGCGGGCCCGCCCAAACTTGACAGCGCTTCCAGGCGGATTTCGGAGGCC	1171
Qy	1535	CCCTGACCTGTGAGAGAGCGCACCTATCTATGGGATAGTGAAGTGGGCTGG	1594
Db	1172	CCCTGGTGTCTGACGATGGCGCATCTTTTGGTGGCATCATCAGTGGGCTGG	1231
Qy	1595	AGTGTGAGAGAGG-----CCAGGGGTCTACACCAAGTTACAAATTCCTGAATTGA	1648
Db	1232	GCTGTGACAGAGGATGTCCTCCGGTGTGTACAAAGTTACCACTACCTAGACTGA	1291
Qy	1649	TCAAAGCCCACT 1661	
Db	1292	TTCTGTGACACAT 1304	

RESULT 15

US-08-883-795A-39  
 ; Sequence 39, Application US/8883795A  
 ; Patent No. 5985607  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Delcuve, Genevieve  
 ; APPLICANT: Awang, Gregor  
 ; TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
 ; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
 ; NUMBER OF SEQUENCES: 39  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: BERESKIN & PARR  
 ; STREET: 40 King Street West  
 ; CITY: Toronto  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: M5H 3Y2  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; FILING DATE: 27-JUN-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Gravelle, Micheline  
 ; REGISTRATION NUMBER: 40,261  
 ; REFERENCE/DOCKET NUMBER: 7841-062  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (416) 364-7311  
 ; TELEFAX: (416) 361-1398  
 ; INFORMATION FOR SEQ ID NO: 39:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1955 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Homo sapiens  
 ; US-08-883-795A-39

Query Match 4.3%; Score 72.2; DB 2; Length 1955;  
 Best Local Similarity 50.5%; Pred. No. 5.4e-12;  
 Matches 340; Conservative 0; Mismatches 288; Indels 45; Gaps 5;

Qy	1028	AGGCCACTTCTGTGGTGGGGCGCTGATCCACCCCTGCTGGTGTCTCACTGTGCCACT	1087
Db	1103	AGCGGTTCTGTGGGGGGCATACTATCAGCTCTCTGATTTCTCTGCGGCCACT	1162
Qy	1088	GCACCGA-----CATAAAAACAGACATCTAAGGTGGTGTAGGGACGAGGACTGA	1141
Db	1163	GCTTCCAGGAGAGGTTTCCGCCCAACACCTGACGGTGATCTTGGCAGAACATACCGG	1222

Search completed: May 25, 2004, 10:46:20  
Job time : 96.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 09:00:39 ; Search time 528.5 Seconds  
(without alignments)  
14472.614 Million cell updates/sec

Title: US-09-912-559-2

Perfect score: 1883  
Sequence: 1 atgttgccaggatgtctga.....aaagtgaagtggtctcttaa 1683

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2953838 seqs, 2272363821 residues

Total number of hits satisfying chosen parameters: 5907676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1683	100.0	1683	9	US-09-912-559-2
2	1683	100.0	1683	16	US-10-391-215-4
3	1681.4	99.9	1683	16	US-10-391-215-2
4	1681.4	99.9	1683	16	US-10-391-215-3
5	1679.8	99.8	1683	9	US-09-912-559-1
6	1679.8	99.8	1683	16	US-10-391-215-1
7	1679.8	99.8	3008	9	US-09-880-107-1668
8	368.4	21.9	428	10	US-09-918-995-8242
9	262.4	15.6	264	13	US-10-425-000-17
10	120.2	7.1	451	9	US-09-864-761-11164
11	117	7.0	117	9	US-09-864-761-27791
12	114.4	6.8	2036	9	US-09-954-456-532
13	114.4	6.8	2036	9	US-09-880-107-1612
14	93.2	5.5	1302	13	US-10-087-192-590

15	93.2	5.5	2299	13	US-10-665-216-10	Sequence 10, Appl
c 16	74.4	4.4	3186778	13	US-10-027-632-174961	Sequence 174961,
c 17	74.4	4.4	3186778	16	US-10-027-632-174961	Sequence 174961,
18	72.8	4.3	614	9	US-09-969-271-6	Sequence 33, Appl
19	72.2	4.3	1689	9	US-09-969-271-6	Sequence 6, Appl
20	72.2	4.3	1726	13	US-10-411-037-25	Sequence 25, Appl
21	72.2	4.3	1726	13	US-10-411-037-25	Sequence 25, Appl
22	72.2	4.3	1726	13	US-10-411-037-25	Sequence 25, Appl
23	72.2	4.3	1726	17	US-10-410-362-25	Sequence 25, Appl
24	72.2	4.3	1726	17	US-10-410-362-25	Sequence 25, Appl
25	72.2	4.3	2461	13	US-10-342-887-541	Sequence 541, App
26	72.2	4.3	2461	13	US-10-342-887-541	Sequence 541, App
27	72.2	4.3	2509	13	US-10-172-118-541	Sequence 7, Appl
28	72.2	4.3	2509	13	US-10-172-118-541	Sequence 5, Appl
29	72.2	4.3	2519	9	US-09-969-271-5	Sequence 540, App
30	72.2	4.3	2519	13	US-10-342-887-540	Sequence 540, App
31	72.2	4.3	2519	13	US-10-342-887-540	Sequence 540, App
32	72.2	4.3	2544	15	US-10-443-701-3	Sequence 3, Appl
33	72.2	4.3	2641	9	US-09-974-298-144	Sequence 144, Appl
34	70.6	4.2	1065	10	US-09-814-353-19992	Sequence 19992, A
35	70.6	4.2	1065	10	US-09-814-353-19992	Sequence 4, Appl
36	70.6	4.2	1065	10	US-09-814-353-19992	Sequence 7, Appl
37	70.6	4.2	1128	10	US-09-987-455-7	Sequence 2, Appl
38	70.6	4.2	1128	10	US-09-987-455-7	Sequence 5, Appl
39	69.2	4.1	1996	16	US-10-401-077-2	Sequence 57, Appl
40	68.6	4.1	8064	15	US-10-004-113-57	Sequence 56, Appl
41	67.8	4.0	8221	12	US-10-152-319A-2167	Sequence 2167, Ap
42	67.8	4.0	505	10	US-09-814-353-14421	Sequence 14421, A
43	67.4	4.0	2160	16	US-10-159-563-132	Sequence 132, App
44	67.4	4.0	2328	15	US-10-247-671-34	Sequence 34, Appl
c 45	66.4	3.9	1078	13	US-10-116-802-163	Sequence 163, App
					US-09-898-837A-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-09-912-559-2  
; Sequence 2, Application US/09912559  
; Patent No. US20020142316A1  
; GENERAL INFORMATION:  
; APPLICANT: ROEMISCH, JUERGEN  
; APPLICANT: STOEHR, HANS-ARNOLD  
; APPLICANT: FEUSSNER, ANNETTE  
; APPLICANT: LANG, WIEGAND  
; APPLICANT: WEIMER, THOMAS  
; APPLICANT: BECKER, MARGRET  
; APPLICANT: NERLICH, CLAUDIA  
; APPLICANT: MUTH-NAUMANN, GUDRUN  
; TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND  
DETECTION METHODS USING SPECIFIC ANTIBODIES  
; FILE REFERENCE: 06478.1457  
; CURRENT APPLICATION NUMBER: US/09/912,559  
; CURRENT FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: DE 100 36 641.4  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: DE 100 50 040.4  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: DE 100 52 319.6  
; PRIOR FILING DATE: 2000-10-21  
; PRIOR APPLICATION NUMBER: DE 101 18 706.8  
; PRIOR FILING DATE: 2001-04-12  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1683  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-912-559-2

Query Match 100.0%; Score 1683; DB 9; Length 1683;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1683; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTGCCAGGATGTCATCTCCATGTTCTCTGTTAATGGCTCTGTTGGAAAGACA 60  
Db 1 ATGTTGCCAGGATGTCATCTCCATGTTCTCTGTTAATGGCTCTGTTGGAAAGACA 60  
QY 61 GCCTGTGGGTTCTCCCTGATGTTCTTTATTGGAAGCTCGACCCAGACTGGACCCCTGAC 120  
Db 61 GCCTGTGGGTTCTCCCTGATGTTCTTTATTGGAAGCTCGACCCAGACTGGACCCCTGAC 120  
QY 121 CAGTATGATTACAGCTACGAGGATTAATATCAGGAAGAGAACACCAAGTAGCACATTACC 180  
Db 121 CAGTATGATTACAGCTACGAGGATTAATATCAGGAAGAGAACACCAAGTAGCACATTACC 180  
QY 181 CATGCTCAGAAATCCTGACTGCTACTACATGAGGACCAAGCTGATCCATGCCAGCCAC 240  
Db 181 CATGCTCAGAAATCCTGACTGCTACTACATGAGGACCAAGCTGATCCATGCCAGCCAC 240  
QY 241 CCCTGTGAACACGCTGGGGAGCTGCTCGTCCATCGGAGCACCTTCAATGCACTGCCCTG 300  
Db 241 CCCTGTGAACACGCTGGGGAGCTGCTCGTCCATCGGAGCACCTTCAATGCACTGCCCTG 300  
QY 301 GCTCCTTCTCTGGGATAGTGTCAAGAGTGCNAATACGTGCAAGGACCAACCATATG 360  
Db 301 GCTCCTTCTCTGGGATAGTGTCAAGAGTGCNAATACGTGCAAGGACCAACCATATG 360  
QY 361 GCGCGGGCCAAATGCTCATTACCCAGAGTCCCTCTACTACCGCTGTGTCTGTAAACAC 420  
Db 361 GCGCGGGCCAAATGCTCATTACCCAGAGTCCCTCTACTACCGCTGTGTCTGTAAACAC 420  
QY 421 CTTTACACAGTCCAGCTGCTCCAGTGGTTCCTGTATGAGGCAAAACCCCTGCCAG 480  
Db 421 CTTTACACAGTCCAGCTGCTCCAGTGGTTCCTGTATGAGGCAAAACCCCTGCCAG 480  
QY 481 AATGGGCTACCTGCTCCCGGCAATAGCGAGATCCAAAGTTCACCTGTGCTGCTCCGAC 540  
Db 481 AATGGGCTACCTGCTCCCGGCAATAGCGAGATCCAAAGTTCACCTGTGCTGCTCCGAC 540  
QY 541 CAGTTCAAGGGGAAATCTGTGAAATAGGTTCTGATGACTGCTATGTGGCGATGGCTAC 600  
Db 541 CAGTTCAAGGGGAAATCTGTGAAATAGGTTCTGATGACTGCTATGTGGCGATGGCTAC 600  
QY 601 TCTTACCGAGGAAATGAATAGACAGTCAACCCAGATCGCTGCTTACTGGAATCC 660  
Db 601 TCTTACCGAGGAAATGAATAGACAGTCAACCCAGATCGCTGCTTACTGGAATCC 660  
QY 661 CACTCTCTTGCAGGAAATTAACAATGTTATGAGGATGCTGAAACCCATGGAT 720  
Db 661 CACTCTCTTGCAGGAAATTAACAATGTTATGAGGATGCTGAAACCCATGGAT 720  
QY 721 GGGGAACACAAATTTCTCAGAAACCCAGATCGGAGCAAAAGCCCTGGTCTTATTA 780  
Db 721 GGGGAACACAAATTTCTCAGAAACCCAGATCGGAGCAAAAGCCCTGGTCTTATTA 780  
QY 781 GTTACCAATGACAGGTAAGTGGAAATCTGTGATGCTCAGCCTGCTCAGCCAGGAC 840  
Db 781 GTTACCAATGACAGGTAAGTGGAAATCTGTGATGCTCAGCCTGCTCAGCCAGGAC 840  
QY 841 GTTSCCTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTCCGGGGTTGACTCC 900  
Db 841 GTTSCCTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTCCGGGGTTGACTCC 900  
QY 901 TGTGGAAGACTGAGTAGCAGAGGAGGAAATCAAGAGATCTATGAGGCTTTAAGAGC 960  
Db 901 TGTGGAAGACTGAGTAGCAGAGGAGGAAATCAAGAGATCTATGAGGCTTTAAGAGC 960  
QY 961 ACAGGGGCAAGCAACCATGGCAGGCTCCCTCAGTCTCGTCTGACCATCTCC 1020  
Db 961 ACAGGGGCAAGCAACCATGGCAGGCTCCCTCAGTCTCGTCTGACCATCTCC 1020  
QY 1021 ATGCCCCAGGCGCACTTCTGTGTGGGCGCTGATCCACCCCTGCTGGTGTCACTGCT 1080  
Db 1021 ATGCCCCAGGCGCACTTCTGTGTGGGCGCTGATCCACCCCTGCTGGTGTCACTGCT 1080

QY 1081 GCCCACTGCACCCAGACATAAAAAACAGACATCTAAAGTGTGCTAGGGGACGAGCCTG 1140  
Db 1081 GCCCACTGCACCCAGACATAAAAAACAGACATCTAAAGTGTGCTAGGGGACGAGCCTG 1140  
QY 1141 AAGAAAGAGAAATTTTCATGAGCAGAGCTTTAGGGTGCAGAAAGATATTCAAGTACAGCCAC 1200  
Db 1141 AAGAAAGAGAAATTTTCATGAGCAGAGCTTTAGGGTGCAGAAAGATATTCAAGTACAGCCAC 1200  
QY 1201 TACAATGAAGAGATGAGATTCCCAATGATTTGCATTTGCTCAAGTTAAAGTCAAGT 1260  
Db 1201 TACAATGAAGAGATGAGATTCCCAATGATTTGCATTTGCTCAAGTTAAAGTCAAGT 1260  
QY 1261 GATGTCACATGTCCTAGAAATCCAAATACGTGAAGACTGTGCTTGTCCCTGATGGTCC 1320  
Db 1261 GATGTCACATGTCCTAGAAATCCAAATACGTGAAGACTGTGCTTGTCCCTGATGGTCC 1320  
QY 1321 TTTCCCTCTGGAGTGTGAGTGCACATCTCTGGTGGGGTGTACAGAAAACAGAAAAGGG 1380  
Db 1321 TTTCCCTCTGGAGTGTGAGTGCACATCTCTGGTGGGGTGTACAGAAAACAGAAAAGGG 1380  
QY 1381 TCCCGCAGCTCCCTGGATGCCAAAGTCAAGTATTGCCAACACTTTGTGCAACTCCCGC 1440  
Db 1381 TCCCGCAGCTCCCTGGATGCCAAAGTCAAGTATTGCCAACACTTTGTGCAACTCCCGC 1440  
QY 1441 CAATCTATGACACATGATGATGACAGTATGATCTGTGAGGAAATCTTCAGAAACCT 1500  
Db 1441 CAATCTATGACACATGATGATGACAGTATGATCTGTGAGGAAATCTTCAGAAACCT 1500  
QY 1501 GGGCAAGACACTCCAGGCTGACTCTGGAGGCCCCCTGCTGTGAGAGGAGCGCAC 1560  
Db 1501 GGGCAAGACACTCCAGGCTGACTCTGGAGGCCCCCTGCTGTGAGAGGAGCGCAC 1560  
QY 1561 TACTACGCTCTATGGATAGTGTGAGTGGGCTGTGAGTGTGAGAGAGGAGCGGCTAC 1620  
Db 1561 TACTACGCTCTATGGATAGTGTGAGTGGGCTGTGAGTGTGAGAGAGGAGCGGCTAC 1620  
QY 1621 ACCCAAGTTACCAATTCCTGAATGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680  
Db 1621 ACCCAAGTTACCAATTCCTGAATGATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680  
QY 1681 TAA 1683  
Db 1681 TAA 1683  
RESULT 2  
US-10-391-215-4  
; Sequence 4, Application US/10391215  
; Publication No. US20040009543A1  
; GENERAL INFORMATION:  
; APPLICANT: KIEHL, STEFAN  
; APPLICANT: WILLEIT, JOHANN  
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF  
; APPLICANT: ROEMISCH, JUERGEN  
; APPLICANT: WEIMER, THOMAS  
; APPLICANT: FEUSSNER, ANNETTE  
; APPLICANT: STOEHR, HANS-ARNOLD  
; APPLICANT: DOERSAM, VOLKER  
; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE  
; FILE REFERENCE: 06478.1457-01  
; CURRENT APPLICATION NUMBER: US/10/391.215  
; CURRENT FILING DATE: 2003-03-19  
; PRIOR APPLICATION NUMBER: 09/912,559  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: DE 100 36 641.4  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: DE 100 50 040.4  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: DE 100 52 319.6  
; PRIOR FILING DATE: 2000-10-21  
; PRIOR APPLICATION NUMBER: DE 101 18 706.8  
; PRIOR FILING DATE: 2001-04-12

***This Page Blank (uspto)***





CURRENT FILING DATE: 2003-03-19  
; PRIOR APPLICATION NUMBER: 09/912,559  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: DE 100 36 641.4  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: DE 100 50 040.4  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: DE 100 52 319.6  
; PRIOR FILING DATE: 2000-10-21  
; PRIOR APPLICATION NUMBER: DE 101 18 706.8  
; PRIOR FILING DATE: 2001-04-12  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1683  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-391-215-2

Query Match 99.9%; Score 1681.4; DB 16; Length 1683;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1682; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	ATGTTTCCAGGATGTCGATCCCATGTTCTGCTGTTAATGCTCTGGTGGAAAGACA	60
DB	1	ATGTTTCCAGGATGTCGATCCCATGTTCTGCTGTTAATGCTCTGGTGGAAAGACA	60
QY	61	GCTGTGGGTTCTCCCTGATGTTCTTTATTCGAAAGCTGGACCCAGACTGGACCCCTGAC	120
DB	61	GCTGTGGGTTCTCCCTGATGTTCTTTATTCGAAAGCTGGACCCAGACTGGACCCCTGAC	120
QY	121	CAGTATGATACAGTACGAGGATTAATACGAGAGAGAACACCAAGTACACATTACC	180
DB	121	CAGTATGATACAGTACGAGGATTAATACGAGAGAGAACACCAAGTACACATTACC	180
QY	181	CATGCTCAGAAATCCTGACTGTTACTACACTGAGAACCAAGCTGATCCATGCCAGCCCAAC	240
DB	181	CATGCTCAGAAATCCTGACTGTTACTACACTGAGAACCAAGCTGATCCATGCCAGCCCAAC	240
QY	241	CCCTGTGAAACAGGTGGGAGTCTGCTGTCATGGGAGCACCTTCACATGCACTGCTCG	300
DB	241	CCCTGTGAAACAGGTGGGAGTCTGCTGTCATGGGAGCACCTTCACATGCACTGCTCG	300
QY	301	GCTCCTTCTCTCGGAATAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAAACCCATGT	360
DB	301	GCTCCTTCTCTCGGAATAGTGTGAGAAAGTGCAGAAATACGTGCAAGGACAAACCCATGT	360
QY	361	GGCGGGGCCAATGTCATTACCCAGATGCTCCCTACTACCGCTGTGTGTAAACAC	420
DB	361	GGCGGGGCCAATGTCATTACCCAGATGCTCCCTACTACCGCTGTGTGTAAACAC	420
QY	421	CCTTACAGAGTCCAGTGTCTCCAGTGTCTCTGTATGCGAGCCAAACCCCTGCCAG	480
DB	421	CCTTACAGAGTCCAGTGTCTCCAGTGTCTCTGTATGCGAGCCAAACCCCTGCCAG	480
QY	481	AATGGGGTACTGCTCCCGCATAAGCGGAGATCCAGTTTCACTGTGCTGTCCGAC	540
DB	481	AATGGGGTACTGCTCCCGCATAAGCGGAGATCCAGTTTCACTGTGCTGTCCGAC	540
QY	541	CAGTTCAAGGGGAATCTGTGAATAGTTTCTGATGACTGCTATGTTGGGATGGCTAC	600
DB	541	CAGTTCAAGGGGAATCTGTGAATAGTTTCTGATGACTGCTATGTTGGGATGGCTAC	600
QY	601	TCTTACCGAGGAAATGAATAGGACACTCAACAGAGATCGTGCCTTTACTGGAATCC	660
DB	601	TCTTACCGAGGAAATGAATAGGACACTCAACAGAGATCGTGCCTTTACTGGAATCC	660
QY	661	CACCTCTCTTGCAGGAGATTAACATGTTTATGGAGGATCTGAAACCCATGGGATT	720
DB	661	CACCTCTCTTGCAGGAGATTAACATGTTTATGGAGGATCTGAAACCCATGGGATT	720
QY	721	GGGGAACCAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGTGTATTATAA	780

DB	721	GGGGAACCAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGGTGTATTATAA	780
QY	781	GTTACCAATGACAAAGTGAAATGGGAATACTGTGATGCTCTCAGCCTGCTCAGCCAGGAC	840
DB	781	GTTACCAATGACAAAGTGAAATGGGAATACTGTGATGCTCTCAGCCTGCTCAGCCAGGAC	840
QY	841	GTTCCCTTACCGAGGAAAGCCCCACTGAGCCATCAACCAAGCTTCGCGGTTTGAATCC	900
DB	841	GTTCCCTTACCGAGGAAAGCCCCACTGAGCCATCAACCAAGCTTCGCGGTTTGAATCC	900
QY	901	TGTGAAAGACTGAGATAGCAGAGAGAGAGATCAAGAGAAATCTATGGAGCTTTAAGAGC	960
DB	901	TGTGAAAGACTGAGATAGCAGAGAGAGAGATCAAGAGAAATCTATGGAGCTTTAAGAGC	960
QY	961	ACGGGGGCAAGCAACCATGGCAGGCGTCCCTCAGTCCTCGCTGCTCTGACCATCTCC	1020
DB	961	ACGGGGGCAAGCAACCATGGCAGGCGTCCCTCAGTCCTCGCTGCTCTGACCATCTCC	1020
QY	1021	ATGCCCCAGGCGCACTTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGTGGTCTCACTGCT	1080
DB	1021	ATGCCCCAGGCGCACTTTCTGTGGTGGGCGCTGATCCACCCCTGCTGGTGGTCTCACTGCT	1080
QY	1081	GCCCACTGCACCGACATAAATAACAGACATCTAAGGTGCTAGGGGACCAAGGACTG	1140
DB	1081	GCCCACTGCACCGACATAAATAACAGACATCTAAGGTGCTAGGGGACCAAGGACTG	1140
QY	1141	AAGAAAGAAATTTCTGATGAGCAGAGCTTTAGGCTGAGAGATATTTCAAGTACAGCCAC	1200
DB	1141	AAGAAAGAAATTTCTGATGAGCAGAGCTTTAGGCTGAGAGATATTTCAAGTACAGCCAC	1200
QY	1201	TACAAATGAAAGATGAGATTCCCAACAATGATATTCGATGCTCAAGTAAAGCCAGTG	1260
DB	1201	TACAAATGAAAGATGAGATTCCCAACAATGATATTCGATGCTCAAGTAAAGCCAGTG	1260
QY	1261	GATGTCCTCTGGGAGTGGCCACATCTCTGGCTGGGCTGTTACAGAAACAGGAAAGGG	1320
DB	1261	GATGTCCTCTGGGAGTGGCCACATCTCTGGCTGGGCTGTTACAGAAACAGGAAAGGG	1320
QY	1321	TTTCCCTCTGGGAGTGGCCACATCTCTGGCTGGGCTGTTACAGAAACAGGAAAGGG	1380
DB	1321	TTTCCCTCTGGGAGTGGCCACATCTCTGGCTGGGCTGTTACAGAAACAGGAAAGGG	1380
QY	1381	TCCCGCAGCTCTGGATGCGCAAGTCAAGCTGATTCGCAACATTTGTGCAACTCCGCG	1440
DB	1381	TCCCGCAGCTCTGGATGCGCAAGTCAAGCTGATTCGCAACATTTGTGCAACTCCGCG	1440
QY	1441	CAACTCTATGACCAATGATGATGACAGATGATGATCTGTGCGAGGAAATCTTCAGAAACCT	1500
DB	1441	CAACTCTATGACCAATGATGATGACAGATGATGATCTGTGCGAGGAAATCTTCAGAAACCT	1500
QY	1501	GGGCAAGACACTGCGCAGGCTGCTCTGGAGGCCCCCTGACCTGTGAGAGAGAGGACCC	1560
DB	1501	GGGCAAGACACTGCGCAGGCTGCTCTGGAGGCCCCCTGACCTGTGAGAGAGAGGACCC	1560
QY	1561	TACTACGCTATGGGATAGTGGGCTGGAGTGTGAGAGAGAGGACCCAGGGGTCTAC	1620
DB	1561	TACTACGCTATGGGATAGTGGGCTGGAGTGTGAGAGAGAGGACCCAGGGGTCTAC	1620
QY	1621	ACCCAGTTACAAATTCCTGAAATGGAATCAAGCCACCATCAAGAGTGAAGTGGCTTC	1680
DB	1621	ACCCAGTTACAAATTCCTGAAATGGAATCAAGCCACCATCAAGAGTGAAGTGGCTTC	1680
QY	1681	TAA 1683	
DB	1681	TAA 1683	

RESULT 4  
US-10-391-215-3  
; Sequence 3, Application US/10391215  
; Publication No. US20040009543A1  
; GENERAL INFORMATION:  
; APPLICANT: KIECHUL, STEFAN

APPLICANT: WILLEIT, JOHANN  
APPLICANT: WIEDERMANN, CHRISTIAN JOSEF  
APPLICANT: ROEMISCH, JUERGEN  
APPLICANT: WEIMER, THOMAS  
APPLICANT: FEUSSNER, ANNETTE  
APPLICANT: STOEHR, HANS-ARNOLD  
APPLICANT: DOERSAM, VOLKER  
TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEINASE  
FILE OF INVENTION: (FSP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS  
FILE REFERENCE: 06478.1457-01  
CURRENT APPLICATION NUMBER: US/10/391.215  
CURRENT FILING DATE: 2003-03-19  
PRIOR APPLICATION NUMBER: 09/912.559  
PRIOR FILING DATE: 2001-07-26  
PRIOR APPLICATION NUMBER: DE 100 36 641.4  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: DE 100 50 040.4  
PRIOR FILING DATE: 2000-10-10  
PRIOR APPLICATION NUMBER: DE 100 52 319.6  
PRIOR FILING DATE: 2000-10-21  
PRIOR APPLICATION NUMBER: DE 101 18 706.8  
PRIOR FILING DATE: 2001-04-12  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 3  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-391-215-3

Query Match 99.9%; Score 1681.4; DB 16; Length 1683;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1682; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGTTGCGCAGAGTCTGATCTCATGTTCTGCTGTTATGCTCTGTTGGGAAAGACA 60  
Db 1 ATGTTGCGCAGAGTCTGATCTCATGTTCTGCTGTTATGCTCTGTTGGGAAAGACA 60

Qy 61 GCCTGTGGGTTCTCCCTGATGCTTTATTGAAAGCCCTGGACCCAGACTGGACCCCTGAC 120  
Db 61 GCCTGTGGGTTCTCCCTGATGCTTTATTGAAAGCCCTGGACCCAGACTGGACCCCTGAC 120

Qy 121 CAGTATGATTACAGTCTCAGAGTATTATCAGGAGAGACACAGTAGCACACTTACC 180  
Db 121 CAGTATGATTACAGTCTCAGAGTATTATCAGGAGAGACACAGTAGCACACTTACC 180

Qy 181 CATGCTGAGATCCCTGACTGTGTTACTACTGAGGACCAAGCTGATCCATGCCGCCAAC 240  
Db 181 CATGCTGAGATCCCTGACTGTGTTACTACTGAGGACCAAGCTGATCCATGCCGCCAAC 240

Qy 241 CCCTGTGACACGGTGGGACTGCTCGTCCATGGGAGACCTTCACATGAGCTGGCTG 300  
Db 241 CCCTGTGACACGGTGGGACTGCTCGTCCATGGGAGACCTTCACATGAGCTGGCTG 300

Qy 301 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAAAATAGCTGCAAGGACAAACCCATGT 360  
Db 301 GCTCCTTTCTCTGGGAATAAGTGTGAGAAAGTGCAAAATAGCTGCAAGGACAAACCCATGT 360

Qy 361 GCGCGGGCCAAATGCTCATACAGAGTCTCCCTACTACTACCGTGTGCTGTGTAACAC 420  
Db 361 GCGCGGGCCAAATGCTCATACAGAGTCTCCCTACTACTACCGTGTGCTGTGTAACAC 420

Qy 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGCCAAACCCCTGCCAG 480  
Db 421 CTTTACAGAGTCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGCCAAACCCCTGCCAG 480

Qy 481 AATGGGCTACTGCTCCCGGATTAAGCGGAGATCCAGTTCACTGCTGCTGCTGCCGAC 540  
Db 481 AATGGGCTACTGCTCCCGGATTAAGCGGAGATCCAGTTCACTGCTGCTGCTGCCGAC 540

Qy 541 CAGTTCAAGGGAAATCTGTGAAATAGTCTGATGATGCTGATGTTGGCGATGGCTAC 600  
Db 541 CAGTTCAAGGGAAATCTGTGAAATAGTCTGATGATGCTGATGTTGGCGATGGCTAC 600

Qy 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCTGCTGCTTTACTGGAATCC 660  
Db 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCTGCTGCTTTACTGGAATCC 660

Qy 661 CACCTCTCTTGCAGGAGAAATTAACAACATGTTTATGAGGATGCTGAAACCATGGGATT 720  
Db 661 CACCTCTCTTGCAGGAGAAATTAACAACATGTTTATGAGGATGCTGAAACCATGGGATT 720

Qy 721 GGGGAACAATTTCTGCAGAAACCCAGATGGGACGAAAGCCCTGCTGCTTTATTAATAA 780  
Db 721 GGGGAACAATTTCTGCAGAAACCCAGATGGGACGAAAGCCCTGCTGCTTTATTAATAA 780

Qy 781 GTTACCAATGACAAGGTGAAATGGGAATACTGTGATGCTCTCAGCCCTGCTCAGCCAGAC 840  
Db 781 GTTACCAATGACAAGGTGAAATGGGAATACTGTGATGCTCTCAGCCCTGCTCAGCCAGAC 840

Qy 841 GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCCGGGTTTACTCC 900  
Db 841 GTTGCTTACCCAGAGAAAGCCCACTGAGCCATCAACCAAGCTTCCGGGTTTACTCC 900

Qy 901 TGTGGAAGACTGAGATAGCAGAGAGGAAAGATCAAGAGATCTATGGAGGCTTTAAGAGC 960  
Db 901 TGTGGAAGACTGAGATAGCAGAGAGGAAAGATCAAGAGATCTATGGAGGCTTTAAGAGC 960

Qy 961 ACGGGGGCAAGCACCATGGCAGGGCTCCCTCCAGTCTCTGCTGCTCTGACCACTCC 1020  
Db 961 ACGGGGGCAAGCACCATGGCAGGGCTCCCTCCAGTCTCTGCTGCTCTGACCACTCC 1020

Qy 1021 ATGCCCCAGGGCCACTTCTGCTGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080  
Db 1021 ATGCCCCAGGGCCACTTCTGCTGGGGCTGATCCACCCCTGCTGGGTGCTCACTGCT 1080

Qy 1081 GCCCACTGACCCGACATAAAAAACAGACATCTAAAGGTGCTGCTAGGGACCCAGACCTG 1140  
Db 1081 GCCCACTGACCCGACATAAAAAACAGACATCTAAAGGTGCTGCTAGGGACCCAGACCTG 1140

Qy 1141 AAGAAAGAGATTTCTAGCAGCAGAGCTTTAGGGTGCAAGATATTCAAGTACAGCCAC 1200  
Db 1141 AAGAAAGAGATTTCTAGCAGCAGAGCTTTAGGGTGCAAGATATTCAAGTACAGCCAC 1200

Qy 1201 TACAATGAAGAGATGAGATTTCCCAATGATATTGCAATGCTCAAGTTAAGCCAGTG 1260  
Db 1201 TACAATGAAGAGATGAGATTTCCCAATGATATTGCAATGCTCAAGTTAAGCCAGTG 1260

Qy 1261 GATGTCATCTGCTCTAGAAATCAAAATACGTGAAGACTGTGCTGCTGCTGCTGCTGCT 1320  
Db 1261 GATGTCATCTGCTCTAGAAATCAAAATACGTGAAGACTGTGCTGCTGCTGCTGCTGCT 1320

Qy 1321 TTTCCCTCTGGGAGTGAGTGCCACATCTCTGCTGGGTGTTTACAGAAAACAGAAAAGGG 1380  
Db 1321 TTTCCCTCTGGGAGTGAGTGCCACATCTCTGCTGGGTGTTTACAGAAAACAGAAAAGGG 1380

Qy 1381 TCCCGCCAGCTCTGATGCCAAAGTCAAGCTGATGCCAACACTTTGTGCACTCCCGC 1440  
Db 1381 TCCCGCCAGCTCTGATGCCAAAGTCAAGCTGATGCCAACACTTTGTGCACTCCCGC 1440

Qy 1441 CAACTCTATGACACACATGATTGATGACAGTATGATCTGTGCAAGAAATCTTCAGAAACCT 1500  
Db 1441 CAACTCTATGACACACATGATTGATGACAGTATGATCTGTGCAAGAAATCTTCAGAAACCT 1500

Qy 1501 GGGCAAGACACTGCGCAGGGTGACTGTGAGGCCCTGCTGAGAGAGAGAGAGAGAGAG 1560  
Db 1501 GGGCAAGACACTGCGCAGGGTGACTGTGAGGCCCTGCTGAGAGAGAGAGAGAGAGAG 1560

Qy 1561 TACTAGCTCTATGGGATGAGCTGGGGCTGGAGTGTGAGAGAGAGAGAGAGAGAGAG 1620  
Db 1561 TACTAGCTCTATGGGATGAGCTGGGGCTGGAGTGTGAGAGAGAGAGAGAGAGAGAG 1620

Qy 1621 ACCCAAGTTACCAAAATTCCTGAAATCGAATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680  
Db 1621 ACCCAAGTTACCAAAATTCCTGAAATCGAATCAAGCCACCATCAAAAGTGAAGTGGCTTC 1680

1691 TAA 1693  
1691 TAA 1693

DB

QY

RESULT 5

US-09-912-559-1

Sequence 1, Application US/09912559

Patent No., US20020142316A1

GENERAL INFORMATION:

APPLICANT: ROEMISCH, JUERGEN

APPLICANT: STOEHR, HANS-ARNOLD

APPLICANT: FEUSSNER, ANNETTE

APPLICANT: LANG, WIEGAND

APPLICANT: WEIMER, THOMAS

APPLICANT: BECKER, MARGRET

APPLICANT: NERLICH, CLAUDIA

APPLICANT: MOTH-NAUMANN, GUDRUN

TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND

TITLE OF INVENTION: DETECTION METHODS USING SPECIFIC ANTIBODIES

FILE REFERENCE: 06478.1457

CURRENT APPLICATION NUMBER: US/09/912.559

CURRENT FILING DATE: 2001-07-26

PRIOR APPLICATION NUMBER: DE 100 36 641.4

PRIOR FILING DATE: 2000-07-26

PRIOR APPLICATION NUMBER: DE 100 50 040.4

PRIOR FILING DATE: 2000-10-10

PRIOR APPLICATION NUMBER: DE 100 52 319.6

PRIOR FILING DATE: 2000-10-21

PRIOR APPLICATION NUMBER: DE 101 18 706.8

PRIOR FILING DATE: 2001-04-12

NUMBER OF SEQ ID NOS: 4

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 1693

TYPE: DNA

ORGANISM: Homo sapiens

US-09-912-559-1

Query Match 99.8%; Score 1679.8; DB 9; Length 1693;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGTTCGCCAGGATGCTCATCTCCATGGTCTTGCTTTTAATGGCTCTGGTGGGAAGACA 60

DB 1 ATGTTCGCCAGGATGCTCATCTCCATGGTCTTGCTTTTAATGGCTCTGGTGGGAAGACA 60

QY 61 GCCTGTGGGTCTCCCTGATGCTTTATTGGAACCTGGACCAGACTGGACCCCTGCAC 120

DB 61 GCCTGTGGGTCTCCCTGATGCTTTATTGGAACCTGGACCAGACTGGACCCCTGCAC 120

QY 121 CAGTATGATTACAGCTACGAGGATTAATAATCAGGAAGAACAACAAGTAGCACCTTACC 180

DB 121 CAGTATGATTACAGCTACGAGGATTAATAATCAGGAAGAACAACAAGTAGCACCTTACC 180

QY 181 CATGCTGAGAACCTCGACTGTTACTACTGAGGACCAAGCTGATCCATGCGACGCCAAC 240

DB 181 CATGCTGAGAACCTCGACTGTTACTACTGAGGACCAAGCTGATCCATGCGACGCCAAC 240

QY 241 CCCTGTGAAACACGGTGGGACTGCCTGCTCCATGGGAGACACCTTCACATGAGCTGCGTG 300

DB 241 CCCTGTGAAACACGGTGGGACTGCCTGCTCCATGGGAGACACCTTCACATGAGCTGCGTG 300

QY 301 GCTCCTTTCTCGGGAATAGTGTGAGAAGTGCAAAAATACGTGCAAGACACACCAATGT 360

DB 301 GCTCCTTTCTCGGGAATAGTGTGAGAAGTGCAAAAATACGTGCAAGACACACCAATGT 360

QY 361 GGCCGGGGCCAATGCTCATTAACAGAGTCCTCCATACCGTGTGTGTGTAACAC 420

DB 361 GGCCGGGGCCAATGCTCATTAACAGAGTCCTCCATACCGTGTGTGTGTAACAC 420

QY 421 CCTTACACAGTCCCAGCTGCTCCCAAGTGGTTCCTGTATGAGGCCCAAAACCCCTGCCAG 480

Qy 1561 TACTAGCTCTATGGGATAGTACGCTGGGCTGGAGTGTGAGAGAGGCGGAGGCTCTAC 1620  
Db 1561 TACTAGCTCTATGGGATAGTACGCTGGGCTGGAGTGTGAGAGAGGCGGAGGCTCTAC 1620  
Qy 1621 ACCCAAGTTACCAAAATTCCTGAATTCGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC 1680  
Db 1621 ACCCAAGTTACCAAAATTCCTGAATTCGATCAAAAGCCACCATCAAAAGTGAAGTGGCTTC 1680  
Qy 1681 TAA 1683  
Db 1681 TAA 1683  
RESULT 6  
US-10-391-215-1  
; Sequence 1, Application US/10391215  
; Publication No. US20040009543A1  
; GENERAL INFORMATION:  
; APPLICANT: KIEHL, STEFAN  
; APPLICANT: WILLEIT, JOHANN  
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF  
; APPLICANT: ROEMISCH, JUERGEN  
; APPLICANT: WEINER, THOMAS  
; APPLICANT: FEUSSNER, ANNETTE  
; APPLICANT: STOEHR, HANS-ARNOLD  
; APPLICANT: DOERSAM, VOLKER  
; TITLE OF INVENTION: (FSAP) I MUTANT OF FACTOR VII ACTIVATING PROTEASE  
; TITLE OF INVENTION: (FSAP) AS RISK FACTOR FOR ARTERIAL THROMBOISIS  
; FILE REFERENCE: 06478.1457-01  
; CURRENT APPLICATION NUMBER: US/10391.215  
; CURRENT FILING DATE: 2003-03-19  
; PRIOR APPLICATION NUMBER: 09/912,559  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: DE 100 36 641.4  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: DE 100 50 040.4  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: DE 100 52 319.6  
; PRIOR FILING DATE: 2000-10-21  
; PRIOR APPLICATION NUMBER: DE 101 18 706.8  
; PRIOR FILING DATE: 2001-04-12  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1683  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-391-215-1  
Query Match 99.8%; Score 1679.8; DB 16; Length 1683;  
Best Local Similarity 99.9%; Pred No. 0;  
Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 ATGTTGCGAGATGCTGATCTCAATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60  
Db 1 ATGTTGCGAGATGCTGATCTCAATGTTCTGCTGTTAATGGCTCTGGTGGGAAGACA 60  
Qy 61 GCCTGTGGTCTCCCTCATGCTCTTTATTGAAAGCCCTGGACCCAGACTGGACCCCTGAC 120  
Db 61 GCCTGTGGTCTCCCTCATGCTCTTTATTGAAAGCCCTGGACCCAGACTGGACCCCTGAC 120  
Qy 121 CAGTATGATTACAGCTACGAGATTATTAATCAGGAAGAAACACCACTAGCACAATTACC 180  
Db 121 CAGTATGATTACAGCTACGAGATTATTAATCAGGAAGAAACACCACTAGCACAATTACC 180  
Qy 181 CATGCTGAGATCTTCATGCTGCTACTGAGGACCAAGCTGATCCATGCCGCCCAAC 240  
Db 181 CATGCTGAGATCTTCATGCTGCTACTGAGGACCAAGCTGATCCATGCCGCCCAAC 240  
Qy 241 CCCTGTGAACACGGTGGGAGCTGCCTCGTCCATGGGAGCACTTCACATGAGCTGCCTG 300  
Db 241 CCCTGTGAACACGGTGGGAGCTGCCTCGTCCATGGGAGCACTTCACATGAGCTGCCTG 300

Qy 301 GCTCCTTTCTCTGGAAATAGTGTGAGAAAGTGCAGAAATAGTGCAGAGGCAACCCCATGT 360  
Db 301 GCTCCTTTCTCTGGAAATAGTGTGAGAAAGTGCAGAAATAGTGCAGAGGCAACCCCATGT 360  
Qy 361 GCGCGGGGCCAATGTCTCATTTACCCAGAGTCTCCCTACTACCCCTGTGTGTAAACAC 420  
Db 361 GCGCGGGGCCAATGTCTCATTTACCCAGAGTCTCCCTACTACCCCTGTGTGTAAACAC 420  
Qy 421 CTTTACACAGTCCGAGCTGCTCCCAAGTGGTTCCTGTATGACGGCCAAACCCCTGCCAG 480  
Db 421 CTTTACACAGTCCGAGCTGCTCCCAAGTGGTTCCTGTATGACGGCCAAACCCCTGCCAG 480  
Qy 481 AATGGGGCTACCTGCTCCCGGCATAAGCGGAGATCCAAAGTTCACCTGTGCTCCTCCGAC 540  
Db 481 AATGGGGCTACCTGCTCCCGGCATAAGCGGAGATCCAAAGTTCACCTGTGCTCCTCCGAC 540  
Qy 541 CAGTTCAAGGGGAAATTCGTGAAATAGTTCCTGATGACTCTCTATGTTGGGATGGCTAC 600  
Db 541 CAGTTCAAGGGGAAATTCGTGAAATAGTTCCTGATGACTCTCTATGTTGGGATGGCTAC 600  
Qy 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCTGCTGCTTTACTTGAACCTCC 660  
Db 601 TCTTACCGAGGAAATGAATAGGACAGTCAACAGCATGCTGCTGCTTTACTTGAACCTCC 660  
Qy 661 CACCTCCTCTTCGAGGAGATTAACAACATGTTTATGAGGATGCTGAAACCCATGGGAT 720  
Db 661 CACCTCCTCTTCGAGGAGATTAACAACATGTTTATGAGGATGCTGAAACCCATGGGAT 720  
Qy 721 GGGGAACACAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGTGTGCTTTATTTAA 780  
Db 721 GGGGAACACAATTTCTGCAGAAACCCAGATCGGACGAAAGCCCTGTGTGCTTTATTTAA 780  
Qy 781 GTTACCAATGACAGGTGAAATGGGAACTGATGATGCTCAGCTGCTCAGCCGAGAC 840  
Db 781 GTTACCAATGACAGGTGAAATGGGAACTGATGATGCTCAGCTGCTCAGCCGAGAC 840  
Qy 841 GTTGCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGCGGTTGACTCC 900  
Db 841 GTTGCTTACCCAGAGGAAAGCCCACTGAGCCATCAACCAAGCTTCGCGGTTGACTCC 900  
Qy 901 TGTGAAAGACTGAGATGACAGAGGAAAGTCAAGAGAACTCTATGAGGCTTTAAGAGC 960  
Db 901 TGTGAAAGACTGAGATGACAGAGGAAAGTCAAGAGAACTCTATGAGGCTTTAAGAGC 960  
Qy 961 ACGGGGGCAAGCACCCTATGGCAGCGTCCCTCCAGTCTCTGCTGCTCTGACCATCTCC 1020  
Db 961 ACGGGGGCAAGCACCCTATGGCAGCGTCCCTCCAGTCTCTGCTGCTCTGACCATCTCC 1020  
Qy 1021 ATGCCCCAGGCGCACTTCTGTGGTGGGCGTGTATCCACCCCTGTGGTGTCTACTGCT 1080  
Db 1021 ATGCCCCAGGCGCACTTCTGTGGTGGGCGTGTATCCACCCCTGTGGTGTCTACTGCT 1080  
Qy 1081 GCCCACTGCACGACATAAAACCCAGACATCTAAAGGTGGTCTAGGGGACCCAGACCTG 1140  
Db 1081 GCCCACTGCACGACATAAAACCCAGACATCTAAAGGTGGTCTAGGGGACCCAGACCTG 1140  
Qy 1141 AAGAAAGAAATTTATGAGCAGAGCTTTAGGGTGCAGAGATATTCAGTACAGCCAC 1200  
Db 1141 AAGAAAGAAATTTATGAGCAGAGCTTTAGGGTGCAGAGATATTCAGTACAGCCAC 1200  
Qy 1201 TACAATGAAGAGATGAGATTCGCCACATGATATTGCTCAAGTTAAAGCCAGTG 1260  
Db 1201 TACAATGAAGAGATGAGATTCGCCACATGATATTGCTCAAGTTAAAGCCAGTG 1260  
Qy 1261 GATGGTCACTGTGCTCAGAAATCCAAATACGTTGAAGACTGTGTGCTGCTGATGGTCC 1320  
Db 1261 GATGGTCACTGTGCTCAGAAATCCAAATACGTTGAAGACTGTGTGCTGCTGATGGTCC 1320  
Qy 1321 TTTCCCTCTGGAGTGAAGTGCACATCTCTGGCTGGGTTTACAGAAACAGAAAGGG 1380  
Db 1321 TTTCCCTCTGGAGTGAAGTGCACATCTCTGGCTGGGTTTACAGAAACAGAAAGGG 1380

QY	1381	TCCGCCAGCTCCTCGGATGCCAAAGTCAAGCTGATTGCCAACACTTTGTGCAACTCCCGC	1440
Db	1381	TCCGCCAGCTCCTCGGATGCCAAAGTCAAGCTGATTGCCAACACTTTGTGCAACTCCCGC	1440
QY	1441	CAACTCTATGACCACATGATTGATGACAGTATGATCTGTGCAGGAAATCTTCAGAAAACCT	1500
Db	1441	CAACTCTATGACCACATGATTGATGACAGTATGATCTGTGCAGGAAATCTTCAGAAAACCT	1500
QY	1501	GGGCAAGACACCTGCCAGGCTGACTGTGGAGGCCCTCGAACCCTGTGAGAAGGACGGCACC	1560
Db	1501	GGGCAAGACACCTGCCAGGCTGACTGTGGAGGCCCTCGAACCCTGTGAGAAGGACGGCACC	1560
QY	1561	TACTACGTCTATGGATAGTGAAGCTGGGGCCTGGAGTGTGAGAAGGACGAGGGGTCTTAC	1620
Db	1561	TACTACGTCTATGGATAGTGAAGCTGGGGCCTGGAGTGTGAGAAGGACGAGGGGTCTTAC	1620
QY	1621	ACCCAAGTTACCAAAATTCCTGAAATGGATCAAAAGCCACATCAAAAGTGAAGTGGCTTC	1680
Db	1621	ACCCAAGTTACCAAAATTCCTGAAATGGATCAAAAGCCACATCAAAAGTGAAGTGGCTTC	1680
QY	1681	TAA 1683	
Db	1681	TAA 1683	
RESULT 7			
US-09-880-107-1668			
; Sequence 1668, Application US/09880107			
; Patent No. US20020142981A1			
; GENERAL INFORMATION:			
; APPLICANT: Horne, Darci T.			
; APPLICANT: Scherf, Uwe			
; APPLICANT: Gene Logic, Inc.			
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer			
; FILE REFERENCE: 44921-5028-WO			
; CURRENT APPLICATION NUMBER: US/09/880,107			
; CURRENT FILING DATE: 2001-06-14			
; PRIOR APPLICATION NUMBER: US 60/211,379			
; PRIOR FILING DATE: 2000-06-14			
; PRIOR APPLICATION NUMBER: US 60/237,054			
; PRIOR FILING DATE: 2000-10-02			
; NUMBER OF SEQ ID NOS: 3950			
; SOFTWARE: Patent In Ver. 2.1			
; SEQ ID NO 1668			
; LENGTH: 3008			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D49742			
US-09-880-107-1668			
Query Match 99.8%; Score 1679.8; DB 9; Length 3008;			
Best Local Similarity 99.9%; Pred. No. 0;			
Matches 1681; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	1	ATGTTTGCCAGAGTCTGTGATCTCATGTTCTGCTGTTAATGCTCTGGTGGAAAGACA	60
Db	97	ATGTTTGCCAGAGTCTGTGATCTCATGTTCTGCTGTTAATGCTCTGGTGGAAAGACA	156
QY	61	GCTGTGGGTTCTCCCTGATGTCCTTAATGAAAGCTGGACCGACAGCTGGACCCCTGAC	120
Db	157	GCTGTGGGTTCTCCCTGATGTCCTTAATGAAAGCTGGACCGACAGCTGGACCCCTGAC	216
QY	121	CAGTATGATTACAGCTACGAGGATTATTAATCAGAAGAACACCAAGTAGCACACTTACC	180
Db	217	CAGTATGATTACAGCTACGAGGATTATTAATCAGAAGAACACCAAGTAGCACACTTACC	276
QY	181	CATGCTGAGAAATCCTGACTGTGATCTACACTGAGGACCAAGCTGATTCATGCGACCCCAAC	240
Db	277	CATGCTGAGAAATCCTGACTGTGATCTACACTGAGGACCAAGCTGATTCATGCGACCCCAAC	336
QY	241	CCCTGGAACACAGCTGGGAGCTGCTCGCTCATGGGAGCACTTTCATGCACTGCTGCTG	300

337	DB	CCCTGTGAACACGGTGGGGAATGCTCTGTCCATGGGACACITTCACATCAGCTGCCTG	396
301	QY	GCTCCTTTCTCTGGGAATPAAGTGTGAGAAAGTGCAAAATACGTGAAGGACAAACCCATGT	360
397	DB	GCTCCTTTCTCTGGGAATPAAGTGTGAGAAAGTGCAAAATACGTGAAGGACAAACCCATGT	456
361	QY	GGCCGGGGCCAAATGCTCATTAACCCAGAGTCTCTCCTACTACCGCTGTGTCTGTAAACAC	420
457	DB	GGCCGGGGCCAAATGCTCATTAACCCAGAGTCTCTCCTACTACCGCTGTGTCTGTAAACAC	516
421	QY	CCTTACACAGTGTCCAGCTGTCTCCAAAGTGTCTGTATGAGGCCAAAACCCCTGCCAG	480
517	DB	CCTTACACAGTGTCCAGCTGTCTCCAAAGTGTCTGTATGAGGCCAAAACCCCTGCCAG	576
481	QY	AATGGGGCTACCTGTCTCCGGCATAAAGCGGAGATCCAAAGTTCACTGTGCTGCCGAC	540
577	DB	AATGGGGCTACCTGTCTCCGGCATAAAGCGGAGATCCAAAGTTCACTGTGCTGCCGAC	636
541	QY	CAGTTCAAGGGGAAATTCGTGAATAGTTCCTGATGACTGCTATGTTGGGATGGCTAC	600
637	DB	CAGTTCAAGGGGAAATTCGTGAATAGTTCCTGATGACTGCTATGTTGGGATGGCTAC	696
601	QY	TCTTACCGAGGAAAATGAATAGGACAGTCAACACGACATGCGTGCTTTACTGGAACTCC	660
697	DB	TCTTACCGAGGAAAATGAATAGGACAGTCAACACGACATGCGTGCTTTACTGGAACTCC	756
661	QY	CACCTCCTCTTCGAGGAAATTTACAACATGTTTATGGAGATGCTGAACCCATGGATT	720
757	DB	CACCTCCTCTTCGAGGAAATTTACAACATGTTTATGGAGATGCTGAACCCATGGATT	816
721	QY	GGGGAACACAATTTCTGCAGAAACCCAGATGCGGACGAAAGCCCTGTCGTTTATTTAA	780
817	DB	GGGGAACACAATTTCTTCGAGAAACCCAGATGCGGACGAAAGCCCTGTCGTTTATTTAA	876
781	QY	GTTTACCAATGACAGGTGAATGGGAACTCTGATGCTCAGCTGCTCAGCCACAGAC	840
877	DB	GTTTACCAATGACAGGTGAATGGGAACTCTGATGCTCAGCTGCTCAGCCACAGAC	936
841	QY	GTTGCCTATCCGAGGAAAGCCCCATGAGCCATCAACCAAGCTTCGGGGTTTGACTCC	900
937	DB	GTTGCCTATCCGAGGAAAGCCCCATGAGCCATCAACCAAGCTTCGGGGTTTGACTCC	996
901	QY	TGTGGAAAGACTGAGATCCAGAGAGAAAGATCAAGAGAACTCTATGGAGGCTTTAAGAGC	960
997	DB	TGTGGAAAGACTGAGATAGCAGAGAGAAAGATCAAGAGAACTCTATGGAGGCTTTAAGAGC	1056
961	QY	ACGCGGGCAAGCACCCATGGCAGGCGTCCCTCCAGTCTCGCTGCTGTGACCATCTCC	1020
1057	DB	ACGCGGGCAAGCACCCATGGCAGGCGTCCCTCCAGTCTCGCTGCTGTGACCATCTCC	1116
1021	QY	ATGCCCGAGGCGACATTCGTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGTCTACTGT	1080
1117	DB	ATGCCCGAGGCGACATTCGTGTGTGGGGCGCTGATCCACCCCTGCTGGGTGTCTACTGT	1176
1081	QY	GCCCATGTCAACGACATATAAAACACAGACATCTTAAGTGGTGTCTAGGGACCGAGACCTG	1140
1177	DB	GCCCATGTCAACGACATATAAAACACAGACATCTTAAGTGGTGTCTAGGGACCGAGACCTG	1236
1141	QY	AAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGCAGAAAGATATTCAAGTACAGCCAC	1200
1237	DB	AAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTGCAGAAAGATATTCAAGTACAGCCAC	1296
1201	QY	TACAATGAAAGAGATGAGATCCCAACATGATATTGCCATTTGCTCAAGTTAAAGCCAGTG	1260
1297	DB	TACAATGAAAGAGATGAGATCCCAACATGATATTGCTCAAGTTAAAGCCAGTG	1356
1261	QY	GATGGTCACTGTGCTCTAGAAATCCAAATACGTAAGACTGTGTGCTGTGCTGATGGTCC	1320
1357	DB	GATGGTCACTGTGCTCTAGAAATCCAAATACGTAAGACTGTGTGCTGTGCTGATGGTCC	1416
1321	QY	TTTCCCTCTGGAGTGAATGCCACATCTCTGGCTGGGTGTTACAGAAACAGAGAAAGGG	1380



;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 11164  
;; LENGTH: 451  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC006097.1  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9

Query Match 7.1%; Score 120.2; DB 9; Length 451;  
Best Local Similarity 97.6%; Pred. No. 1.2e-28;  
Matches 122; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 324 TCAGAAAGTCAAAATACGTGCAAGGACCAACCCATGTGCGCGGGGCAATGTCTCATTTAC 393  
Db 451 TCCTACAGTCAAAATACGTGCAAGGACCAACCCATGTGCGCGGGGCAATGTCTCATTTAC 392  
Qy 384 CCAGAGTCTCCCTACTACCGCTGTGTCTGTAAACACCCCTTACACAGGTCCTCCAGCTGCTC 443  
Db 391 CCAGAGTCTCCCTACTACCGCTGTGTCTGTAAACACCCCTTACACAGGTCCTCCAGCTGCTC 332  
Qy 444 CCAAG 448  
Db 331 CCAAG 327

RESULT 11  
US-09-864-761-27791/c  
; Sequence 27791, Application US/09864761  
; Patent No. US20020048763A1

;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
;; FILE REFERENCE: Acomica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 27791  
;; LENGTH: 117  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC006097.1  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9  
;; OTHER INFORMATION: NT HIT: S83182.1, EVALUATE 8.00e-60  
;; OTHER INFORMATION: SWISSPROT HIT: Q25464, EVALUATE 1.00e-03  
;; OTHER INFORMATION: EST\_HUMAN HIT: T68666.1, EVALUATE 7.00e-55  
US-09-864-761-27791

Query Match 7.0%; Score 117; DB 9; Length 117;  
Best Local Similarity 100.0%; Pred. No. 6.2e-28;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 332 TGCAAAATACGTGCAAGGACCAACCCATGTGCGCGGGGCAATGTCTCATTTACCCAGAGTC 391  
Db 117 TGCAAAATACGTGCAAGGACCAACCCATGTGCGCGGGGCAATGTCTCATTTACCCAGAGTC 58



QY	392	CTCCCTACTACCGCTGTGCTGTGTAACACACCCCTTACACAGGTCCCGAGCTGCTCCCAAG	448
Db	57	CTCCCTACTACCGCTGTGCTGTGTAACACACCCCTTACACAGGTCCCGAGCTGCTCCCAAG	1
RESULT 12			
US-09-954-456-552			
; Sequence 552, Application US/09954456			
; Patent No. US20020115057A1			
; GENERAL INFORMATION:			
; APPLICANT: Young, Paul			
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can			
; FILE REFERENCE: 589290-76			
; CURRENT APPLICATION NUMBER: US/09/954,456			
; CURRENT FILING DATE: 2001-09-18			
; PRIOR APPLICATION NUMBER: US/60/233,617			
; PRIOR FILING DATE: 2000-09-18			
; PRIOR APPLICATION NUMBER: US/60/234,052			
; PRIOR FILING DATE: 2000-09-20			
; PRIOR APPLICATION NUMBER: US/60/234,923			
; PRIOR FILING DATE: 2000-09-25			
; PRIOR APPLICATION NUMBER: US/60/235,134			
; PRIOR FILING DATE: 2000-09-25			
; PRIOR APPLICATION NUMBER: US/60/235,637			
; PRIOR FILING DATE: 2000-09-26			
; PRIOR APPLICATION NUMBER: US/60/235,638			
; PRIOR FILING DATE: 2000-09-26			
; PRIOR APPLICATION NUMBER: US/60/235,711			
; PRIOR FILING DATE: 2000-09-27			
; PRIOR APPLICATION NUMBER: US/60/235,720			
; PRIOR FILING DATE: 2000-09-27			
; PRIOR APPLICATION NUMBER: US/60/235,840			
; PRIOR FILING DATE: 2000-09-27			
; PRIOR APPLICATION NUMBER: US/60/235,863			
; PRIOR FILING DATE: 2000-09-27			
; NUMBER OF SEQ ID NOS: 2276			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 552			
; LENGTH: 2036			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-954-456-552			
Query Match 6.8%; Score 114.4; DB 9; Length 2036;			
Best Local Similarity 49.6%; Pred. No. 2.5e-26;			
Matches 605; Conservative 0; Mismatches 561; Indels 54; Gaps 10;			
QY	469	AACCCCTGCCAGATGGGCTACTCTCTCCGGCATAGCGGAGATCCAGTTCACTGT	528
Db	745	AGCCCTTGGCTGAACGGGGGCACCTGCCCTGATCGTGGCCACCGGACCCCGTGTGT	804
QY	529	GCCTGTCCCACCAGTTCAAGGGGAAATTCTGTGAATAGGTTCTGATGA---CTGCTAT	585
Db	805	GCCTGCCCCACCAGCTTCGCTGGACGGCTCTGCAACATCGAGCCTGATGAGCCTGCTTC	864
QY	586	GTGGGGCATGGTACTCTTACGAGGGAAATGAATAGGACAGTCAACACAGTACGTGTC	645
Db	865	TTGGGGAACGGCACTGGGTAACCGTGGCGTGCCAGCACCTCAGCCTCGGSCCTCAGCTGC	924
QY	646	CTTTACTGGAATCCCACTCCCTCTTTCAGAGGAATTACAAATGTTATGGAGGATGCT	705
Db	925	CTGGCTTGGAACTCCGATCTGCTCTACGAGAGCTGCACTGGACTCCGTGGGCGCGG	984
QY	706	GAACACCATCGGATTTGGGGAACACAAATTTCTGCAGAAACCCAGATCGGACGAAAGCCC	765
Db	985	GCCTGTGTGGGCTGGGCCCCCATGCTACTTGGCGGAATCCGGACAATGACGAGAGGCC	1044
QY	766	TGTTGCTTTTAAAGTTACCAATGACAAAGTGAATGGGAATCTGTGATGTTCTCAGCC	825
Db	1045	TGTTGCTACTGT---GGTGAAGAGACAGCGGCTCTCTCTGGGAGTACTGCCCGCTTGGAGGCC	1101
QY	826	TGCTCA3CCCGAGACGCTTGCTCTACCCAGAGAAAGCCCACTCAGGCCATCAACCAAGCTT	885

RESULT 13

```

RESULI 13
US-09-880-107-1512
; Sequence 1612, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14

```

; PRIOR APPLICATION NUMBER: US 60/211,379  
 ; PRIOR FILING DATE: 2000-06-14  
 ; PRIOR APPLICATION NUMBER: US 60/237,054  
 ; PRIOR FILING DATE: 2000-10-02  
 ; NUMBER OF SEQ ID NOS: 3950  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1612  
 ; LENGTH: 2036  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D14012  
 ; US-09-880-107-1612

Query Match 6.8%; Score 114.4; DB 9; Length 2036;  
 Best Local Similarity 49.6%; Pred. No. 2.5e-26;  
 Matches 605; Conservative 0; Mismatches 561; Indels 54; Gaps 10;  
 QY 469 AACCCCTGCGAGATGGGGTACCTGCTCCGGCATAACGGAGATCCAGTTCACTGT 528  
 DB 745 AGCCCTTGCGTGAACGGGGGACCTGCCACCTGATCGTGGCCACCGGACCCGTGT 804  
 QY 529 GCCTGCTCCGACACAGTTCAAGGGGAAATCTGTGAATAGTTCTGATGA---CTGCTAT 585  
 DB 805 GCCTGCCACACAGCTTCGCTGACGGCTCTGAACATCGACCTGATGAGCGCTGCTTC 864  
 QY 586 GTTGGCGATGGCTACTCTTACCGAGGAAATGAATAGACAGTCAACGAGATCGGTGC 645  
 DB 865 TTGGGAACGGCACTGGTACCGTGGCGAGCCTCGACCTCGGGCCCTCAGCTGC 924  
 QY 646 CTTTACTGGAATCCACCTCTCTTGCAGAGAAATTACAATGTTTATGAGGATGCT 705  
 DB 925 CTGGCTGGAATCCGATCTGCTTACAGAGAGCTGCACTGGACTCGTGGCGCGCG 984  
 QY 706 GAAACCATGGGATTTGGGAACAAATTTCTGAGAAACCCAGATCGGAGCAAAAGGCC 765  
 DB 985 GCCTGTGGGCTGGGGCCCTGCTACTGCGGAATCCGGCAATCAGCAGAGAGGCC 1044  
 QY 766 TGTGCTTTTAAAGTTACCAATGACAGGTGAATGGAATACCTGATCTCTCAGCC 825  
 DB 1045 TGTGCTACGT---GGTGAAGACAGCGGCTCTCTGGAGTACTGCGCCTGGAGGCC 1101  
 QY 826 TGTCTAGCCAGACGTTGCTTACCAGAGAAAGCCCACTGAGCCATCAACCAAGCTT 885  
 DB 1102 TGGCAATCCCT-----CACCAGAGTCCAACTGTCCCGGATCTCCTGGCGACCTG 1152  
 QY 886 CCGGGTTTCACTCTCTGGAAGACTGAGATGACAGAGGAAGATCAAGAGATCTAT 945  
 DB 1153 CTGAGCCAGCTCCCGGGGCGCCAGGCTGCGGAGAGGCAAGAGAGGAGCTTC 1212  
 QY 946 GGAGGCTTTAAGAGCAGCGCGGCAAGCAACCCATGCGAGCGTCCCTCCAGTCTCGTG 1005  
 DB 1213 CTGCGGCCAGTATCATCGCGGCTCTCTCTGCTGCGGCTCGCACCCCTG----- 1265  
 QY 1006 CTTCTGACCATCTCCATGCCCCAGGGCCACTTCTGTGGGGCGTGTATCCACCCCTGC 1065  
 DB 1266 --GCTGGCGCCCATCTACATCGGGGACAGCTTCTGCGCGGGAGCGCTGTCTCACACCTGC 1323  
 QY 1066 TGGGTGCTACTGTGCGCCACTGCGACATATAAACCAGACATCTAA-----AGGTG 1119  
 DB 1324 TGGGTGCTGCGCGGCCACTGCTTCTCCACAGCCCCCCCCAGGACAGGCTCTCCGTG 1383  
 QY 1120 GTCTAGGAGCAGAGCTGGAAGAAGAATTTATGAGAGAGCTTTAGGTGTCAG 1179  
 DB 1384 GTCTGGCGCAGACATTTCTTCAACCGGACGACGAGCGTGAACGAGACCTTCGCGATCGAG 1443  
 QY 1180 AAGATATTCAAGTACAGCCACTACAATGAAGAGATGAGATTTCCCAACATATATGCA 1239  
 DB 1444 AAGTACATCCGTACACCTGTACTCGGTGTTCAACCCAG---CGACCCAGCCTCGTC 1500  
 QY 1240 TTGCTCAAGTTAAGCCAGTGGATGCTCACTGTCTCTAGATCCAAATACGTGAAGACT 1299  
 DB 1501 CTGATCCGGCTGAAGAAAGGGGACCGCTGTGCCACACGCTCGAGTTCGTGACGCC 1560

RESULT 14

US-10-087-192-590  
 ; Sequence 590, Application US/10087192  
 ; Publication No. US20020182586A1  
 ; GENERAL INFORMATION: David W.  
 ; APPLICANT: Morris, David W.  
 ; APPLICANT: Engelhard, Eric K.  
 ; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
 ; FILE OF INVENTION: CANCER  
 ; FILE REFERENCE: 529452000122  
 ; CURRENT APPLICATION NUMBER: US/10/087,192  
 ; CURRENT FILING DATE: 2002-03-01  
 ; PRIOR APPLICATION NUMBER: US 09/747,377  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: US 09/798,586  
 ; PRIOR FILING DATE: 2001-03-02  
 ; NUMBER OF SEQ ID NOS: 2059  
 ; SOFTWARE: PastSeq for Windows Version 4.0  
 ; SEQ ID NO 590  
 ; TYPE: DNA  
 ; ORGANISM: Mus musculus  
 ; US-10-087-192-590

Query Match 5.5%; Score 93.2; DB 13; Length 1302;  
 Best Local Similarity 49.7%; Pred. No. 2.1e-19;  
 Matches 586; Conservative 0; Mismatches 523; Indels 69; Gaps 11;

QY 526 TGTGCTGTCCCGACAGTTCAAGGGGAAATTTCTGAAATAGGT---TCTGATGATGC 582  
 DB 154 TGCAGCTGCCCAAGGAAATTCAGGGGAGCACTGTGAGATAGATGATCAAAAACCTGC 213  
 QY 583 TATGTTGGCATGGCTACTCTTACCGAGGAAATGAATAGGACAGTCAACCCAGCATGCG 642  
 DB 214 TATCATGAAATGTGTACTCTTACCGAGGAAAGCCAACTGATACCAAGTGGGCC 273  
 QY 643 TGCCTTTACTGGAATCCCACTCTCTTGCAGGAAATTTACAACATGTTTATGGAGAT 702  
 DB 274 TGCCTGGCTGGAATGGCGCTGTCTTTCAGAAACCTTACAATGCCACAGACCTGAT 333  
 QY 703 GCTGAACCCATGGATGGGAAACAAATTTCTGCAAGAAACCCAGATCGGCGGAAAG 762  
 DB 334 GCTATTAGCTAGGCTGGGGGAAACAAATTTCTGCAAGAAACCCCTGCAACCAAGCGCA 393

763 CCTGGTCTTTATTAAGTTACCAATGACAAGGTGAATGGAATACTGTGATGTCTCA 822  
Db |||||  
394 CCTGGTCTATGTGCAGATGGCTTAAGCAGTT-----TGCCAA 435  
Qy |||||  
823 GCCTGCTCAGCCAGGACGTTGCTACCCAGAGAAAGCCCACTGAGCCATCAACCAAG 882  
Db |||||  
436 GAATGCATGTTGCATGAC--TGCTCTCTTAGCAAAAAGCCTTCTTGCTGTGAGACCAAC 493  
Qy |||||  
883 CTTCCGGGGTTGACTCTCTGTGMAAAGTGTGATAGCAGAGAGGAAGATCAAGAAATC 942  
Db |||||  
494 AAGCTTCCAGTGTGGCC-----AGAAGCTCTAAGGCCCGCCTTTAAGATT 540  
Qy |||||  
943 TATGAGGCTTTAAGACGCGGGCAAGCACCATTGGCAGGCGTCCCTCCAGTCTCTCG 1002  
Db |||||  
541 GTTGGGGAGAAATTCATGAGTGTGAGAACAGCCCTGGTTCGAGCCAT-----CTAC 594  
Qy |||||  
1003 CTGCTCTGACCATCTCCATGCCCCAGGGCCACTTCTGTGTGGGGGGCTGTATCCACCC 1062  
Db |||||  
595 CAGAAGAACACAGGAGGAAGTCTCTCCCTCTTAAATGTGTGGAGTCTCATCAGTCT 654  
Qy |||||  
1063 TGCCTGGTGTCTACTGCTGCCACTGCAAC-----GACATAAAACACGACATCTAAAG 1116  
Db |||||  
655 TGCCTGGTGGCCAGTGGCCACACTGCTTCAATCACTCCAAAGAGAAACTAGCTT 714  
Qy |||||  
1117 GTGTGCTAGGGACACAGGACCTGAAGAAAGAAAGATTTATGAGCAGAGCTTTAGGGTG 1176  
Db |||||  
715 GTCTACCTGGTCTAGTCCGAGAGAGCTCTATAATCTGTGAGAGATGAAGTTGAGGTG 774  
Qy |||||  
1177 CAGAGATATTCAGTACAGCCACTAGATGAAGAGATGAGATTCCTCCACATGATTT 1236  
Db |||||  
775 GAGCAGCTCATCTTGCACGAATATCTACAGGAGACAGCCTGGCTACCATATGATTT 834  
Qy |||||  
1237 GCATTGCTCAAGTTAAAGCCAGTGGTGTCTCTGATGATCTCTAGAAATCAATAGTGAAG 1296  
Db |||||  
835 GCCTTGTGTAAGATACGTACAGCAGCGGCAATGTGCACAGCCATCCAGTCCATACAG 894  
Qy |||||  
1297 ACTGTGCTTGCCTGATGGTCTT-----TCCCTCTGGAGTGTGCTCCATCTCT 1350  
Db |||||  
895 ACCATCTGCTGCCCCCAAGCTTACTGATGCTCCGTTTGGTTCCAGACTGTGAGATCACT 954  
Qy |||||  
1351 GCCTGGGT---GTTACAGAAACAGGAAAGGGTCCCGCCAGCTCTCTGGA---TGCCAAA 1404  
Db |||||  
955 GCTTTGAAAGAGTCTGAAAGTACTATCTCTATCCAAAGAACCTGMAATGTCGTC 1014  
Qy |||||  
1405 GTCAAGCTGATGCAACACATTTGTGCACTCCCGCCCACTCTATGACCATGATGAT 1464  
Db |||||  
1015 GTAAAGCTTGTCTCATGAACAGTGTATGCAGGCCCACTACTATGGCTCTGAAATTAAT 1074  
Qy |||||  
1465 GACATGATGCTGTGCAGGAAATCTTCAGAAACCTGGGCAAGACACCTGCCAGGCTGAC 1524  
Db |||||  
1075 TATAAATGCTGTGCTGCGGACCCAGTGTGAAAC---AGATTCTGCAAGGGCGAT 1131  
Qy |||||  
1525 TCTGAGGCCCTTCACTGTGAGAGGACCGCACCTACTAGTCTATGGAATAGTGAGC 1584  
Db |||||  
1132 TCTGAGGACCGCTTATCTGTAACATCGAAGGCCGCCCACTCTGAGTGGGATTTGAGC 1191  
Qy |||||  
1585 TGGGCCCTGGAGTGG-----AGAGAGGCCGCGGTCTACCCCAAGTTTACCAATTC 1638  
Db |||||  
1192 TGGGCCCGAGGATGTGAGAGAAACAAAGCCCGGTGTCTACAGAGGGTCTCACACTTC 1251  
Qy |||||  
1639 CTGAATTGGATCAAGCCCACTCAAAAGTGAAGTGG 1676  
Db |||||  
1252 CTGGACTGGATTCAATCCACATTCGAGAGAGAAAGG 1289

## RESULT 15

US-10-665-216-10

; Sequence 10, Application US/10665216

; Publication No. US20040043957A1

; GENERAL INFORMATION:

; APPLICANT: Brenda F. Baker

; APPLICANT: Susan M. Freier

APPLICANT: Andrew T. Watt  
TITLE OF INVENTION: ANTISENSE MODULATION OF UROKINASE PLASMINOGEN ACTIVATOR EXPRESSIO  
FILE REFERENCE: RTS-0188  
CURRENT APPLICATION NUMBER: US/10/665,216  
CURRENT FILING DATE: 2003-09-19  
PRIOR APPLICATION NUMBER: US/09/821,972  
PRIOR FILING DATE: 2001-03-30  
NUMBER OF SEQ ID NOS: 168  
SEQ ID NO 10  
LENGTH: 2299  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (59)...(1360)  
US-10-665-216-10

Query Match 5.5%; Score 93.2; DB 13; Length 2299;

Best Local Similarity 49.7%; Pred. No. 2.9e-19; Indels 59; Gaps 11;

Matches 586; Conservative 0; Mismatches 523;

Qy 526 TGTGCTGTCTCCGACCAAGTTCAGAGGGGAAATTTCTGTGAAATAGGT---TCTGATGACTGC 582  
Db 212 TGCAGCTGCCCAAGGAAATTCAGGGGGAGCACCTGTGAGATAGATGCATCAAAAACCTGC 271  
Qy 583 TATGTTGGGATGGCTACTCTTACCGAGGGAATATGATAGCAGTCAACAGCATGG 642  
Db 272 TATCATGGAATGGTACTCTTACCGAGGAAAGGCCAACCTGATACCAAAGTCCGCC 331  
Qy 643 TGCCTTTACTGAACTCCACCTCTCTTGCAGGAGAAATTACAACTGTTTATGGAGAT 702  
Db 332 TGCCTGGCTGGAATGCCCTGCTCTCTCAGAAACCTTCAATGCCACAGACTGAT 391  
Qy 703 GTGAAACCCATGGGATTTGGGGAACACAAATTTCTCAGAAACCCAGATGCGGACGAAAG 762  
Db 392 GCTATTAGCTAGGCTTGGGGAACACAAATTTACTCAGGAAACCCCTGACCAACGAGCGA 451  
Qy 763 CCTGTGCTTTTATAAGTTTACCAATGACAAGGTGAATGGAATACTGTGATGTCTCA 822  
Db 452 CCTGTGCTATGTGCAAGTTGGCTTAAGCAGTT-----TGTCAA 493  
Qy 823 GCCTGCTAGCCAGAGCTTCCCTACCCAGAGGAAAGCCCACTGAGGCATCAACCAAG 882  
Db 494 GAATGCATGTCATGAC--TGCTCTCTTAGCAAAAAGCCTTCTCTGCTGTGAGACCAAC 551  
Qy 883 CTTCCGGGTTTGAATCTCCTGTGAAAGACTGAGATAGCAGAGAGGAAGATCAAGAGAATC 942  
Db 552 AAGGCTTCCAGTGTGGCC-----AGAAGGCTTAAGGCCCGCTTTAAGATT 598  
Qy 943 TATGAGGCTTTAAGAGCACGCGGGCAAGCACCCATGSCAGGCGTCCCTCCAGTCTCG 1002  
Db 599 GTTGGGGAGAAATTCATCAGGTGGAGAACCAAGCCCTGTTCCAGCCAT-----CTAC 652  
Qy 1003 CTGCTCTGACCATCTCCATGCCCCAGGCCACTTCTGTGTGGGGCGCTGATCCACCC 1062  
Db 653 CAGAAGAACAAAGGAGGAAGTCTCTCCCTCTTTAAATGTGTGGGAGTCTCATCAGTCT 712  
Qy 1063 TGTGCTGTCTCACTGCTGCCCACTGCACC-----GACATAAAAACCAAGACATCTAAAG 1116  
Db 713 TGTGCTGTGGCAGTGGCCGACACACTGCTTCAATCAACTCCCAAGAGGAAACTACGTT 772  
Qy 1117 GTGCTGCTAGGGACCCAGGACCTTGAAGAAAGAAATTTTCATGAGCAGAGCTTTAGGGTG 1176  
Db 773 GTCTACCTGGGTCACTGGAAGGAGAGTCTCTTAATCTCTGGAGAGATGAAGTTTGGGGTG 832  
Qy 1177 CAGAAGATATTCAGTACAGCCACTACAATGAAGAGATGAGATTCCTCCCAATGATATT 1236  
Db 833 GAGCAGCTCATCTTGCACGNAATACTACAGGGAAGACAGCCTGGCTTACCATATGATATT 892  
Qy 1237 GCATTGCTCAAGTTAAAGCCAGTGGGATGTGCTACTGTGCTCTAGAAATCCAAATACGTGAAG 1296  
Db 893 GCCTTGTGGAAGATAGTATCCAGCAGCGGCCAATGTGACAGCCATCCAGGTCCATACAG 952

QY	1297	ACTGTGTGCTTGCCTGAGTGGTCCCTT-----TCCTCTGGAGTGGTGGCCACATCTCT	1350
Db	953	ACCATCTGCTGCCCCCAAGGTTTACTGATGCTCCGTTTGGTTTCAGACTGTGAGATCACT	1012
QY	1351	GGCTGGGT---GTTACAGAAACAGGAAAGGTCGCCAGCTCCTTGA---TGCCAAA	1404
Db	1013	GGCTTTGGAAGAGTCTGAAAGTGAATCTCTATCCAAAGAACTGAAATGTCCGTC	1072
QY	1405	GTCAAGCTGATTGCCAACACTTTGTGCAACTCCCGCAACTCTATGACCAATGATTGAT	1464
Db	1073	GTAAGCTTGTCTTCTCATGAACAGTGTATGCAGCCCCCACTACTATGGCTCTGAAATTAAT	1132
QY	1465	GACAGTATGATCTGTGCAGGAATCTTCAGAAACCTGGGCAAGACACCTGCCAGGTGAC	1524
Db	1133	TATAAAATGCTGTGCTGGGACCCAGAGTGGAAAC---AGATTCTGCAAGGCGAT	1189
QY	1525	TCTGGAGGCCCTGACCTGTGAGAAAGGACGGCACCTACTACGTCTATGGGATAGTGAGC	1584
Db	1190	TCTGGAGGACCGCTTATCTGTAACATCGAAGGCCGCCCAACTCTGAGTGGGATTGTGAGC	1249
QY	1585	TGGGGCTGAGTGTG-----AGAGAGGCCAGGGGTCTACACCCCAAGTTACCAAATTC	1638
Db	1250	TGGGGCGGAGGATGTGAGAGAAACAGAGCCCGGTCTTACACGAGGCTCTCACATTC	1309
QY	1639	CTGAATTGGATCAAAGCCACCATCAAAGTGAAGTGG	1676
Db	1310	CTGAGCTGGATTCAATCCCAATTGGAGAGAGAAAGG	1347

Search completed: May 25, 2004, 13:22:39  
Job time : 537 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

```

n: May 24, 2004, 09:39:54 ; Search time 54 Seconds
    (without alignments)
    2930.122 Million cell updates/sec

```

Title: US-09-912-559-3  
Perfect score: 3154  
Sequence: 1 MFARMSDLHYLLLMALVGT.....TQVTKFLNWKATIKSESGF 560

Scoring table: BLOSUM62  
Gapop 10.0 ; Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database : A_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003a.*
7: Geneseq2003bs.*
8: Geneseq2004s.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query			DB	ID	Description
		Match	Length	%			
1	3154	100.0	560	3	AAB42484	Aab42484 Human ORF	
2	3154	100.0	560	5	AAO17144	Aao17144 Human blo	
3	3143	99.7	560	5	AAO17145	Aao17145 Human blo	
4	833.5	26.4	655	2	AAR53962	Aar53962 Hepatocyt	
5	832.5	26.4	655	2	AAR98197	Aar98197 Human hep	
6	812.5	25.8	441	1	AAP70643	Aap70643 Modified	
7	812.5	25.8	473	1	AAP70642	Aap70642 Modified	
8	807.5	25.6	689	7	ADE07850	Ade07850 Novel pro	
9	802	25.4	439	2	AAR09260	Aar09260 t-PA vari	
10	798.5	25.3	434	2	AAR13922	Aar13922 Delta (92-	
11	797	25.3	439	2	AAR09261	Aar09261 t-PA vari	
12	795	25.2	439	2	AAR09258	Aar09258 t-PA vari	
13	794	25.2	439	2	AAR09259	Aar09259 t-PA vari	
14	793	25.1	623	1	AAP17142	Aap17142 tle(277)l	
15	792	25.1	623	1	AAP17140	Aap17140 t-PA wic	
16	786	24.9	527	2	AAR70858	Aar70858 Human t-P	
17	786	24.9	562	2	AAR41545	Aar41545 Modified	
18	785	24.9	518	2	AAR74684	Aar74684 t-PA mute	
19	785	24.9	522	2	AAR13917	Aar13917 Delta (46	
20	784	24.9	527	2	AAR70893	Aar70893 Human t-P	
21	783	24.8	521	2	AAR74678	Aar74678 t-PA mute	
22	782.5	24.8	522	2	AAR09257	Aar09257 t-PA vari	
23	782	24.8	483	2	AAR74683	Aar74683 t-PA mute	
24	781.5	24.8	483	2	AAR09252	Aar09252 t-PA vari	
25	781	24.8	519	2	AAR74687	Aar74687 t-PA mute	

## ALIGNMENTS

## RESULT 1

AAB42484  
 ID AAB42484 standard; protein; 560 AA.

AAC AAB42484:

DT 08-FEB-2001 (first entry)

Human OREF ORF2248 polypeptide sequence SEQ ID NO:4496.

Human; open reading frame, ORFX; detection; cytostatic; hepatotropic; KW  
KW vulnary; antipsoriatic; antiparkinsonian; necrotropic; neuroprotective; KW  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac; KW  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; KW  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory; KW  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid; KW  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension; KW  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease; AIDS; KW  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; KW  
KW cholesterol ester storage; systemic lupus erythematosus; infection; KW  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma; KW  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; KW  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation; KW  
KW thrombosis; contraceptive.

xx Homo sapiens.

PN WO200058473-A2.

XX  
PD  
05-OCT-2000.

31-MAR-2000: 2000WC-US008621.

XX  
PR 31-MAR-1999: 99US-0127607P.

02-APR-1999; 5305-0127833E  
05-APR-1999; 99US-0127728P

PR 30-MAR-2000; 2000US-00340783.  
YY

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M;  
yy

DR WPI; 2000-602362/57.  
DR N. DODD: AXC76693

XX  
E  
[New York]

PT useful for treati

XX PS Claim 11; Page 3681-3683; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnery;

CC antiproliferative; antiparkinsonian; necrotic; neuroprotective; osteopathic;

CC anticancer; antidiabetic; immunosuppressant; immunostimulant;

CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;

CC dermatological; immunosuppressive; antinflammatory; antibacterial;

CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The

CC sequences can be used for determining the presence of or predisposition

CC to or preventing or treating pathological conditions associated with an

CC ORFX-associated disorder. The nucleic acids can be used to express ORFX

CC proteins in gene therapy vectors. The proteins and nucleic acids may be

CC used to treat cancers, proliferative disorders, neurodegenerative

CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,

CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester

CC storage, systemic lupus erythematosus, severe combined immunodeficiency

CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune

CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and

CC cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to

CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX SQ Sequence 560 AA;

Query Match 100.0%; Score 3154; DB 3; Length 560;  
Best Local Similarity 100.0%; Pred. No. 7.2e-181; Indels 0; Gaps 0;  
Matches 560; Conservative 0; Mismatches 0;

Qy 1 MFARMSDLHVLMLVKTACGFSLSLLESLEDPDTPDQYDYSYEDYNOENTSTLT 60

Db 1 MFARMSDLHVLMLVKTACGFSLSLLESLEDPDTPDQYDYSYEDYNOENTSTLT 60

Qy 61 HAENPDWYTTEDQADPCPNCEHGGDCLVHGSGFTCSCLAPSGNKCKQVNTCKDNPC 120

Db 61 HAENPDWYTTEDQADPCPNCEHGGDCLVHGSGFTCSCLAPSGNKCKQVNTCKDNPC 120

Qy 121 GRGQCLITQSPYRYVCVKCHPYTGSCSQVVPVCRPNPCNGATCSRHRKRSKFTACPD 180

Db 121 GRGQCLITQSPYRYVCVKCHPYTGSCSQVVPVCRPNPCNGATCSRHRKRSKFTACPD 180

Qy 181 QFKGKFCIEGSDDCYVGDGYSGYRGKMRVTNQHACLWNHLLLOENYMFEDAEETHGI 240

Db 181 QFKGKFCIEGSDDCYVGDGYSGYRGKMRVTNQHACLWNHLLLOENYMFEDAEETHGI 240

Qy 241 GEHNFRCNPDADKRPWCFIKVTNDKWKVEYCDVSACSADQVAYPEESPTSTKLPGFDS 300

Db 241 GEHNFRCNPDADKRPWCFIKVTNDKWKVEYCDVSACSADQVAYPEESPTSTKLPGFDS 300

Qy 301 CGKTEIABRKIKRIYGGFKSTAGKHPWQASLQSSPLTISMPQGHFCGGLIHPCWYLT 360

Db 301 CGKTEIABRKIKRIYGGFKSTAGKHPWQASLQSSPLTISMPQGHFCGGLIHPCWYLT 360

Qy 361 AHCTDIKTRHLKVLGDQDLKKEEFHESFVEKIFKYSYNERDEIPHNDIALLLKPV 420

Db 361 AHCTDIKTRHLKVLGDQDLKKEEFHESFVEKIFKYSYNERDEIPHNDIALLLKPV 420

Qy 421 DGHCALESKYVTKVCLPDGSPFSGSECHISGWVETGKGRQLLDKXVLIANTLNSR 480

Db 421 DGHCALESKYVTKVCLPDGSPFSGSECHISGWVETGKGRQLLDKXVLIANTLNSR 480

Qy 481 QLYDHMIDSMICAGNLQKPGQDTQCGSGGGLICEKDGTVYVYGVISWGLECCRPVY 540

Db 481 QLYDHMIDSMICAGNLQKPGQDTQCGSGGGLICEKDGTVYVYGVISWGLECCRPVY 540

Qy 541 TQVTKFLNWKATIKSESGF 560

Db 541 TQVTKFLNWKATIKSESGF 560

RESULT 2  
AAO17144

ID AAO17144 standard; protein; 560 AA.

XX AC AAO17144;

XX DT 13-JUN-2002 (first entry)

XX DE Human blood coagulation factor VII activating protease.

XX KW Human; blood coagulation factor VII activating protease; FSAP;

XX KW single-chain plasminogen activator; bleeding disorder; haematological;

XX KW haemostatic.

XX OS Homo sapiens.

XX PN EP1182258-A1.

XX PD 27-FEB-2002.

XX PF 05-JUL-2001; 2001EP-00115691.

XX PR 26-JUL-2000; 2000DE-01036641.

XX PR 10-OCT-2000; 2000DE-01050040.

XX PR 21-OCT-2000; 2000DE-01052319.

XX PR 12-APR-2001; 2001DE-01018706.

XX PA (AVET ) AVENTIS BEHRING GMBH.

XX PI Roemisch J, Stoehr H, Feussner A, Lang W, Weimer T, Becker M;

XX PI Kerlich C, Muth-Naumann G;

XX DR WPI; 2002-270939/32.

XX DR N-PSDB; AAL45696.

XX PT New nucleic acid encoding mutant factor 7 activating protease, useful for

XX PT diagnosis, treatment and prevention of coagulation disorders, also

XX PT related protein and antibodies.

XX PS Claim 4; Page 17-19; 27pp; German.

XX The present invention relates to a mutant of the DNA sequence encoding

XX the protease (FSAP) that activates blood coagulation factor VII (FVII)

XX and single-chain plasminogen activator, where at least one of the base

XX changes G to C at nucleotide 1177 and G to A at nucleotide 1601 is

XX present. The mutant sequences can be used in the treatment and prevention

XX of bleeding disorders associated with inherited or acquired defects of

XX blood coagulation factors V, VIII, IX, X, XI, XII, von Willebrand's

XX factor and/or with antibodies against any of these proteins. The present

XX sequence is the human FSAP protein

XX SQ Sequence 560 AA;

Query Match 100.0%; Score 3154; DB 5; Length 560;  
Best Local Similarity 100.0%; Pred. No. 7.2e-181; Indels 0; Gaps 0;  
Matches 560; Conservative 0; Mismatches 0;

Qy 1 MFARMSDLHVLMLVKTACGFSLSLLESLEDPDTPDQYDYSYEDYNOENTSTLT 60

Db 1 MFARMSDLHVLMLVKTACGFSLSLLESLEDPDTPDQYDYSYEDYNOENTSTLT 60

Qy 61 HAENPDWYTTEDQADPCPNCEHGGDCLVHGSGFTCSCLAPSGNKCKQVNTCKDNPC 120

Db 61 HAENPDWYTTEDQADPCPNCEHGGDCLVHGSGFTCSCLAPSGNKCKQVNTCKDNPC 120

Qy 121 GRGQCLITQSPYRYVCVKCHPYTGSCSQVVPVCRPNPCNGATCSRHRKRSKFTACPD 180

Db 121 GRGQCLITQSPYRYVCVKCHPYTGSCSQVVPVCRPNPCNGATCSRHRKRSKFTACPD 180

Qy 181 QFKGKFCIEGSDDCYVGDGYSGYRGKMRVTNQHACLWNHLLLOENYMFEDAEETHGI 240

Db 181 QFKGKFCIEGSDDCYVGDGYSGYRGKMRVTNQHACLWNHLLLOENYMFEDAEETHGI 240

Qy 241 GEHNFRCNPDADKRPWCFIKVTNDKWKVEYCDVSACSADQVAYPEESPTSTKLPGFDS 300

241 GEHNFCRNPDAKEKWCPIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300  
301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCGALIHPCWVLT 360  
301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCGALIHPCWVLT 360  
361 AHCTDIKTRHLKVLGDDQLKKEEFHEQSFRVEKIFKYSHYNERDEIPHNDIALLKLPV 420  
361 AHCTDIKTRHLKVLGDDQLKKEEFHEQSFRVEKIFKYSHYNERDEIPHNDIALLKLPV 420  
421 DGHCALESKYVKTVCPLPGSPSGSECHISGNGVTETGKGRQLLDKAVKLIANTLNSR 480  
421 DGHCALESKYVKTVCPLPGSPSGSECHISGNGVTETGKGRQLLDKAVKLIANTLNSR 480  
481 QLYDHMIDDSMICAGNLQKPGQDTCCQDSGGPLTCEKDGTYVYGVISWGLECKRPGVY 540  
481 QLYDHMIDDSMICAGNLQKPGQDTCCQDSGGPLTCEKDGTYVYGVISWGLECKRPGVY 540  
541 TQVTKFLNWKATIKSES GF 560  
541 TQVTKFLNWKATIKSES GF 560

RESULT 3  
AAO17145  
ID AAO17145 standard; protein; 560 AA.  
XX  
AC AAO17145;  
DT 13-JUN-2002 (first entry)  
XX  
DE Human blood coagulation factor VII activating protease mutant.  
XX  
KW Human; blood coagulation factor VII activating protease; FSAP;  
KW single-chain plasminogen activator; bleeding disorder; haematological;  
KW haemostatic; mutant; mutein.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN EP1182258-A1.  
XX  
PD 27-FEB-2002.  
XX  
PF 05-JUL-2001; 2001EP-00115691.  
XX  
PR 26-JUL-2000; 2000DE-01036641.  
PR 10-OCT-2000; 2000DE-01050040.  
PR 21-OCT-2000; 2000DE-01052319.  
PR 12-APR-2001; 2001DE-01018706.  
XX  
PA (AVET ) AVENTIS BEHRING GMBH.  
XX  
PI Roemisch J, Stoehr H, Feussner A, Lang W, Weimer T, Becker M;  
PI Nerlich C, Muth-Naumann G;  
XX  
DR WPI; 2002-270939/32.  
DR N-PSDB; AAL45697.  
XX  
PT New nucleic acid encoding mutant factor 7 activating protease, useful for  
PT diagnosis, treatment and prevention of coagulation disorders, also  
PT related protein and antibodies.  
XX  
PS Disclosure; Page 20-22; 27pp; German.  
XX  
CC The present invention relates to a mutant of the DNA sequence encoding  
CC the protease (FSAP) that activates blood coagulation factor VII (FVII)  
CC and single-chain plasminogen activator, where at least one of the base  
CC changes G to C at nucleotide 1177 and G to A at nucleotide 1601 is  
CC present. The mutant sequences can be used in the treatment and prevention  
CC of bleeding disorders associated with inherited or acquired defects of  
CC blood coagulation factors V, VII, IX, X, XI, XII, von Willebrand's  
CC factor and/or with antibodies against any of these proteins. The present

CC sequence is the mutant human FSAP protein  
XX  
SQ Sequence 560 AA;  
Query Match 99.7%; Score 3143; DB 5; Length 560;  
Best Local Similarity 99.6%; Pred. No. 3.3e-180;  
Matches 559; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MFARMSDLHLVLLMALVGTACGFSLSLLESLLDPDPTDQYDYSYEDYNQEBNTSSTLT 60  
DB 1 MFARMSDLHLVLLMALVGTACGFSLSLLESLLDPDPTDQYDYSYEDYNQEBNTSSTLT 60  
QY 61 HAENPDWYTTEDQADPCQNPCEHGGDCLVHSGTFTCSCLAPFSGNCKQVQNTCKDNPC 120  
DB 61 HAENPDWYTTEDQADPCQNPCEHGGDCLVHSGTFTCSCLAPFSGNCKQVQNTCKDNPC 120  
QY 121 GRQCILITQSPPYRCVCKHPYTGPSQVQVPCRENPCQNGATCSRHKRSKFTCACPD 180  
DB 121 GRQCILITQSPPYRCVCKHPYTGPSQVQVPCRENPCQNGATCSRHKRSKFTCACPD 180  
QY 181 QFKGKFCIGSDDCYVGDGYSYRGKNRTVNOHACLYNWSHLLLOENYMFMEDEATHGI 240  
DB 181 QFKGKFCIGSDDCYVGDGYSYRGKNRTVNOHACLYNWSHLLLOENYMFMEDEATHGI 240  
QY 241 GEHNFCRNPDAKEKWCPIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300  
DB 241 GEHNFCRNPDAKEKWCPIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300  
QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCGALIHPCWVLT 360  
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCGALIHPCWVLT 360  
QY 361 AHCTDIKTRHLKVLGDDQLKKEEFHEQSFRVEKIFKYSHYNERDEIPHNDIALLKLPV 420  
DB 361 AHCTDIKTRHLKVLGDDQLKKEEFHEQSFRVEKIFKYSHYNERDEIPHNDIALLKLPV 420  
QY 421 DGHCALESKYVKTVCPLPGSPSGSECHISGNGVTETGKGRQLLDKAVKLIANTLNSR 480  
DB 421 DGHCALESKYVKTVCPLPGSPSGSECHISGNGVTETGKGRQLLDKAVKLIANTLNSR 480  
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCCQDSGGPLTCEKDGTYVYGVISWGLECKRPGVY 540  
DB 481 QLYDHMIDDSMICAGNLQKPGQDTCCQDSGGPLTCEKDGTYVYGVISWGLECKRPGVY 540  
QY 541 TQVTKFLNWKATIKSES GF 560  
DB 541 TQVTKFLNWKATIKSES GF 560

RESULT 4  
AAR53962  
ID AAR53962 standard; protein; 655 AA.  
XX  
AC AAR53962;  
XX  
DT 25-MAR-2003 (revised)  
DT 06-JAN-1995 (first entry)  
XX  
DE Hepatocyte growth factor converting protease.  
XX  
KW Hepatocyte growth factor; protease; cleavage; active; inactive;  
KW precursor.  
XX  
OS Homo sapiens.  
XX  
PN EP596524-A2.  
XX  
PD 11-MAY-1994.  
XX  
PF 05-NOV-1993; 93EP-00117988.  
XX  
PR 05-NOV-1992; 92JP-00296133.  
PR 20-NOV-1992; 92JP-00312234.



```
PR 20-NOV-1992; 92JP-00312242.
XX PA (MITU ) MITSUBISHI KASEI CORP.
XX Shimomura T, Yamada K, Morimoto Y, Kitamura N, Miyazawa K;
XX WPI; 1994-152921/19.
XX N-PSDB; AAQ63951.
XX Hepatocyte growth factor converting protease and precursor and gene
PT encoding them - for producing active two chain HGF from inactive single
PT chain HGF.
XX Claim 12; Page 21-24; 30pp; English.
XX Hepatocyte growth factor converting protease is capable of converting
CC inactive single chain hepatocyte growth factor (HGF) into active two
CC chain HGF by cleavage at a specific site. (Updated on 25-MAR-2003 to
CC correct PN field.)
XX PS Sequence 655 AA;
XX
XX Query Match 26.4%; Score 833.5; DB 2; Length 655;
XX Best Local Similarity 33.6%; Pred. No. 6.5e-42;
XX Matches 192; Conservative 89; Mismatches 220; Indels 71; Gaps 21;
QY 31 ESLDPDWTDPQDYSDYEDYNEQENTSTLTHAENPDW-----YYTEDQA----- 74
Db 99 QALTEDGRPCPFPRYGRMLHACTSEGSARHK---WCATTHNYDRDRAGYCVETATPPP 155
QY 75 -----DPCQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKC--QKQVNTCK-----D 117
Db 156 GGPAALDPCASGCLNGGSCSNTQDPOSYHCSCRAFTGKDCGTEKCFDETREYLEGGD 215
QY 118 NPGCGGCLITQPPYRVCYKHPYTGPSQSVV--PVCRPNPCQNGATCSRHRKRSKFT 175
Db 216 RWARVRQCHVEQ-----CSC---FGGRTWCEGTRHTACLSPLNGGTCCHLIVATGTV 266
QY 176 CACPDQKGFCEIGSD--CYVGDGYSYRGKMRNTVNOHACLYNWSHLLLOENYMFMD 234
Db 267 CACPPPGFAGRLCNIEPDERCFNGTGYRGVASTSASGJSCGLAWSNLDLYQELHVDVGA 326
QY 235 AETHGIGEHNFRCNPDADEKPFKVTNDKWKVEYCDVSACSA--QDVAYPESPTSP 292
Db 327 AALLGLGPHAYCRNPDNDRPWCYV-VKDSALSWEYCRUEACBSLTVQLSPDLLATLPE 385
QY 293 TKLPFGFDSCKTEIAERKIK-RIYGFKSTAGKHPQWASLQSSLPDTISMPOGHCEGAL 351
Db 386 PASFGRCACGRHKRTFLRPRIIGSSSLFGSHPMLAIYIG-----DSFCAGSL 436
QY 352 IHPCWVLTAAHCTDIKTRH--LKVLGDQDLKKEEFHQSFVEKIFKYSHYNERDEIPH 409
Db 437 VHTCWVSAACFSHSPRDSVSVLGQHFENRTDVTQTFG:EKYIPYTLISVFNPSDH 496
QY 410 NDIALLLKLPVDGHALESKYKTVCLPD--GSFPSSGECHISGWG-VTEGKG-BRQLL 465
Db 497 -DLVLRLLKKKGDRCATRSQFVQICLPEPGSTFPAGHKQIAGWGHLDENVSGYSSSLR 555
QY 466 DAKVLLIANTLNCNSRQLYDHMDISMCAGNLQKPGODTCQDGGGGLTCCKDGTYYVG 525
Db 556 EALVPLVADHKCSSFEVIGADISFNMUJCAGVFDC-SDACQDGGGGLACRNGVAILYG 614
QY 526 IVSWGLECGK--RPGVYTVQTKFLNWKATIK 555
Db 615 IISWGDGCGRLHKGTVYRVANYVDINDRIR 646
RESULT 5
AAR89197
ID AAR89197 standard; protein; 655 AA.
XX AC
XX AAR89197;
XX
```

```
DT 08-AUG-1996 (first entry)
XX Human hepatocellular growth factor single chain precursor protein.
XX Mature protein; inactive; single chain; hepatocellular growth factor;
XX liver; human; precursor; dimerisation; double chain; serine protease;
XX hepatitis.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Peptide 356..655
XX FT /note="mature peptide"
XX JPO8027026-A.
XX 30-JAN-1996.
XX 22-JUL-1994; 94JP-00171207.
XX 22-JUL-1994; 94JP-00171207.
XX (MITU ) MITSUBISHI CHEM CORP.
XX WPI; 1996-136206/14.
XX New preventative and therapeutic cpds contg. a 34 kD serine protease -
XX useful for treatment of liver diseases e.g. hepatitis.
XX Claim 4; Page 6-8; 8pp; Japanese.
XX This is the amino acid sequence of the precursor protein of the inactive
XX single chain form of a hepatocellular growth factor. The mature protein
XX (AAR89196) has a mol. wt. of around 34 kD and is derived from the 96 kD
XX precursor protein. The mature protein corresp. to residues 356-655 of the
XX precursor protein. The inactive single chain form of the growth factor is
XX activated by dimerisation of the mature protein. The active protein is a
XX serine protease which can be used for the treatment of hepatitis
XX
XX Query Match 26.4%; Score 832.5; DB 2; Length 655;
XX Best Local Similarity 33.2%; Pred. No. 7.4e-42;
XX Matches 192; Conservative 87; Mismatches 216; Indels 83; Gaps 21;
QY 31 ESLDPDWTDPQDYSDYEDYNEQENTSTLTHAENPDW-----YYTEDQA----- 74
Db 99 QALTEDGRPCPFPRYGRMLHACTSEGSARHK---WCATTHNYDRDRAGYCVETATPPP 155
QY 75 -----DPCQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKCQKQVNTCKDNPCGQOCL 126
Db 156 GGPAALDPCASGCLNGGSCSNTQDPOSYHCSCRAFTG-----KD--CGTEKCF 203
QY 127 ITQSPPYV-----RCVCKHPYTGSCSQVW--PVCRPNPCQNGATCSRHK 169
Db 204 DETREYLEGGDRWARVRQGHVEQCEC---FGGRTWCEGTRHTACLSPLNGGTCCHLIV 260
QY 170 RRSKFTCACPQKGFCEIGSD--CYVGDGYSYRGKMRNTVNOHACLYNWSHLLLOENY 228
Db 261 ATGTTVCACPGFAGRLCNIEPDERCFNGTGYRGVASTSASGLSLAWSNLDLYQELH 320
QY 229 NMFMEDAETHGIGEHNFRCNPDADEKPFKVTNDKWKVEYCDVSACSA--QDVAYPEE 286
Db 321 VDSVGAALLGLGPHAYCRNPDNDRPWCYV-VKDSALSWEYCRUEACBSLTVQLSPDLL 379
QY 287 SPTEPSTKLPQFDSCKTEIAERKIK-RIYGFKSTAGKHPQWASLQSSLPDTISMPOGH 345
Db 380 LATLEPASPRQACGRHKRTFLRPRIIGSSSLFGSHPMLAIYIG-----DS 430
QY 346 FCGGALIHPCWVLTAAHCTDIKTRH--LKVLGDQDLKKEEFHQSFVEKIFKYSHYNE 403
Db 431 FCAGSLVHTCWVSAACFSHSPRDSVSVLGQHFENRTDVTQTFG:EKYIPYTLISV 490
```

QY 404 RDEIPHNDIALKLKPVGDGHCALESKYVKTVCPLD--GSFPGBGEBCHISGNG-VTETGK 460  
 Db 491 FNPSPH-DLVLRLKKKGDRCATRSQFVQPCLPBPGSTFFAGHKCQIAGHGLDENVSG 549  
 QY 461 -SRQLLDKAVKLIANTLNSRQLYDHMIDDSMICAGNLQKPGQDTCCQSDSGGPTCEKDG 519  
 Db 550 YSSSLREALVPLVADHKCSSPEVYGADISPNMLCAGYFDCK-SDACQSDSGGPLACEKNG 608  
 QY 520 TYYVYGIYSWGIECGK--RPGVYTVTKFLNWKATIK 555  
 Db 609 VAYLYGISWGDCGRLLKPGVYTVRYVYVDMINDRIR 646

## RESULT 6

AAP70643

ID AAP70643 standard; protein; 441 AA.

XX AC AAP70643;

XX DT 25-MAR-2003 (revised)

XX DT 10-APR-1991 (first entry)

XX DE Modified tissue plasminogen activator.

XX KW Tissue plasminogen activator; kringle domain; embolism; thrombosis;

XX KW stroke.

XX OS Homo sapiens.

XX PN AU8661804-A.

XX PD 05-MAR-1987.

XX PF 25-AUG-1986; 86AU-00061804.

XX PR 26-AUG-1985; 85US-00769298.

XX PA (ELIL ) LILLY &amp; CO ELI.

XX PI Bang NU, Little SP, Schoner BE, Weigel BJ;

XX DR WPI; 1987-108842/16.

XX DR N-PSDB; AAN70992.

XX PT New modified tissue plasminogen activator - with new encoding DNA, new

XX PT DNA expression vector, useful for treating vascular disorders, e.g.,

XX PT pulmonary embolism arterial thrombosis.

XX PS Claim 9; Page 122; 164pp; English.

XX CC The modified t-PA has all/part of the kringle domains of native t-PA

XX CC removed. The t-PA has functional properties superior to those of native t

XX CC -PA. It retains fibrin binding properties and interacts more slowly and

XX CC inefficiently with plasminogen activator inhibitor(s) compared to native

XX CC t-PA. It is obtd. in large ants. from a prokaryotic host. Modified t-PA

XX CC is used for treating vascular disorders, eg deep vein thrombosis,

XX CC pulmonary embolism, peripheral arterial thrombosis, disseminated

XX CC intravascular coagulation, emboli from the heart or peripheral arteries,

XX CC acute myocardial infarction, thrombotic strokes or fibrin deposits

XX CC associated with invasive cancers. t-PA is used at a dosage of 250000 to

XX CC 500000 units at a loading dose or in a deep vein thrombosis-pulmonary

XX CC embolism, or 250000-750000 units over 30-90 mins. in acute myocardial

XX CC infarction. (Updated on 25-MAR-2003 to correct PA field.)

XX SQ Sequence 441 AA;

SQ Query Match 25.8%; Score 812.5; DB 1; Length 441;

Best Local Similarity 37.4%; Pred. NO. 8e-41;

Matches 177; Conservative 70; Mismatches 163; Indels 59; Gaps 14;

QY 112 QNTCKNPGCRGQCLITQSPYPYR-----CVCKHPYTGPSQVVPV--CRPNPC 159

Db 5 QVICRDE--XTQWIVQCHQSWLRPVLRSNRVEYWCN---SGRACHSVPKVSCSEPRC 58

QY 160 QNGATCSRHKRSKFTACPDQKGFCEIGSDDCVVDGYSYRGKMRNTVQHCACLYWN 219  
 Db 59 FNGGTQQALYFSDFYCQCPGEGFAGKSCIEGNSDCYFGNGSAVRGTHSLTESGASCLPWN 118  
 QY 220 SHLLLENYMFNEDAEHGEHNEPCRPDADDERKWCIFKVTNDKVKWEYCDVSNCSAQ 279  
 Db 119 SMILIGKVTAQNPQAALGLGKHNYCRNPDDGAKWCHV-LKNRRLTWYCDVPSCS-- 175  
 QY 280 DVAYPEESPTEPSTKLPDGFDCGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSILPLTI 339  
 Db 176 -----TCGLRYSQPF-RIKGGLFADIASHPWQAIFAK--HRR 212  
 QY 340 SMPQGHFGGALHPQWLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSFVEKIFK 397  
 Db 213 SPGERFLCGGILISSCWILSAACFOERPPPHLTVILGRTYRVWPGEQEKEVEKYIV 272  
 QY 398 YSHYNERDEIPHNDIALKLKPVGDGHCALESKYVKTVCPLD--DGSPFSGSECHISGNGVT 455  
 Db 273 HKEFD--DDTYDNDIALQLKSDSSRCAQESSVVRVCLPPADLQLPDWTECELSGYGXH 330  
 QY 456 ETGKG--SRQLLDKAVKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----ODTCQGD 508  
 Db 331 EALSPFYSERLKEAHVRLYPSRSCTSQHLINRTVTDNMLCAGDTRSGGPPQANLHDCQGD 390  
 QY 509 SGGPLCEKDGTYVYVYGIYSWGIECGK--RPGVYTVTKFLNWKATIK 555  
 Db 391 SGGPLVCLNDGRMTLVGLISWGLCGCKQKQVPGVYTKVTNYLDWIRDNR 439

## RESULT 7

AAP70642

ID AAP70642 standard; protein; 473 AA.

XX AC AAP70642;

XX DT 25-MAR-2003 (revised)

XX DT 10-APR-1991 (first entry)

XX DE Modified tissue plasminogen activator.

XX KW Tissue plasminogen activator; kringle domain; embolism; thrombosis;

XX KW stroke.

XX OS Homo sapiens.

XX PN AU8661804-A.

XX PD 05-MAR-1987.

XX PF 25-AUG-1986; 86AU-00061804.

XX PR 26-AUG-1985; 85US-00769298.

XX PA (ELIL ) LILLY &amp; CO ELI.

XX PI Bang NU, Little SP, Schoner BE, Weigel BJ;

XX DR WPI; 1987-108842/16.

XX DR N-PSDB; AAN70991.

XX PT New modified tissue plasminogen activator - with new encoding DNA, new

XX PT DNA expression vector, useful for treating vascular disorders, e.g.,

XX PT pulmonary embolism arterial thrombosis.

XX PS Claim 9; Page 122; 164pp; English.

XX CC The modified t-PA has all/part of the kringle domains of native t-PA

XX CC removed. The t-PA has functional properties superior to those of native t

XX CC -PA. It retains fibrin binding properties and interacts more slowly and

XX CC inefficiently with plasminogen activator inhibitor(s) compared to native

XX CC t-PA. Obtd. in large ants. from a prokaryotic host. Modified t-PA is used

XX CC for treating vascular disorders, eg deep vein thrombosis, pulmonary



QY 543 VTKFLNWKATIK 555  
 DB 668 VANYVDWINDQIR 680

RESULT 9  
 AAR09260  
 ID AAR09260 standard; protein; 439 AA.  
 XX  
 AC AAR09260;  
 DT 24-FEB-1993 (first entry)  
 DE t-PA variant d92-179, N184D, I210R, G211A, K212R, V213R, T252R, F305H.  
 XX  
 KW Tissue plasminogen activator; zymogen; clot; plasma; plasmin.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 91..92  
 FT /note= "amino acids 92-179 of the wild-type, mature t-PA  
 FT have been deleted"  
 FT Misc-difference 96  
 FT /label= substitution  
 FT /note= "N184 of the wild-type, mature t-PA has been  
 FT substituted for D"  
 FT Misc-difference 122  
 FT /label= substitution  
 FT /note= "I210 of the wild-type, mature t-PA has been  
 FT substituted for R"  
 FT Misc-difference 123  
 FT /label= substitution  
 FT /note= "G211 of the wild-type, mature t-PA has been  
 FT substituted for A"  
 FT Misc-difference 124  
 FT /label= substitution  
 FT /note= "K212 of the wild-type, mature t-PA has been  
 FT substituted for R"  
 FT Misc-difference 125  
 FT /label= substitution  
 FT /note= "V213 of the wild-type, mature t-PA has been  
 FT substituted for R"  
 FT Misc-difference 184  
 FT /label= substitution  
 FT /note= "T252 of the wild-type, mature t-PA has been  
 FT substituted for R"  
 FT Misc-difference 217  
 FT /label= substitution  
 FT /note= "F305 of the wild-type, mature t-PA has been  
 FT substituted for H"  
 FT  
 PN WO9002798-A.  
 XX  
 XX 22-MAR-1990.  
 XX  
 PF 02-SEP-1988; 88US-00240856.  
 XX  
 PR 02-SEP-1988; 88US-00240856.  
 PR 24-JUL-1989; 89US-00384608.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Anderson S, Bennett WE, Botstein D, Higgins DL, Paoni NF;  
 PI Zoller M;  
 XX  
 XX WPI; 1990-115987/15.  
 XX  
 XX Modified tissue plasminogen activator - activated only when proximate to  
 PT plasmin at site of clot and not systemically.  
 PT  
 XX Claim 17; Page 44; 63pp; English.  
 PS

XX The t-PA variants represented in AAR09215-63 (and the T252R or N184S  
 CC analogues thereof or combinations thereof) and AAR09266-83 have their  
 CC fibrinolytic activity turned off when generally in the plasma and  
 CC activated when proximate to plasmin at the site of the clot thus  
 CC providing specific localized clot therapy  
 XX  
 SQ Sequence 439 AA;  
 Query Match 25.4%; Score 802; DB 2; Length 439;  
 Best Local Similarity 37.7%; Pred No. 3.4e-40;  
 Matches 177; Conservative 68; Mismatches 165; Indels 60; Gaps 15;  
 QY 112 QNTCKDNPCGRQCCLITQSPYYR-----CVCXHPYTPGSCSQVVPV--CRPNPC 159  
 DB 3 QVICRDE---KTQMIYQQHQSRLRFLRSNRVEYWCN---SGRAQCHSVVPKCSSEPRC 56  
 QY 160 QNGATCSRHKRSKFTCACPDQFKGKFCBIGS-DDCYVDGDYSGYRGKMRRTVQNHACLYW 218  
 DB 57 FNGGTCQALYFSDFCVQCPEGFAGKCEIDTRATCYFGDSAYRGTHSLTESGASCLPW 116  
 QY 219 NSHLLAQENYMFMDAETHGIGENFNCPADSKPWCIFKVTNDKVKWEYCDVSACSA 278  
 DB 117 NSMILRARRYTAQNPSPAQALGLGKHYCRNPDGDAKPCWCHV-LKNRLRWKEYCDVPSCS- 174  
 QY 279 QDVAYPEESPTPESTKLPGFSDCGKTEIAERIKRIYGGFKSTAGKHPQASLQSLPLT 338  
 DB 175 -----TCGLRQYSPQF-RIKGSLFADIASHPQAAIFAK--HR 210  
 QY 339 ISMPQGHFCGGALIHPCWVLTAHCTD--IKTRHLKVLGDQDLKKEEFHEQSFRVEKIF 396  
 DB 211 RSPGHERLCCGILISSCWILSAHCFQERPPPHLLTVILGRTVYVVPGEQKFEVEKYI 270  
 QY 397 KYSHYNERDEIPHNDIALLKLPVDGHCALSKYKVTVCPL--DGSFPGSGSECHISGMGV 454  
 DB 271 VHKEFD--DITYDNDIALQLKSDSSRCAQESSVVRIVCLPPADLQLPDWTCELSGYK 328  
 QY 455 TETGKG--SRQLLDKVKLIANTLCKNSRLQYDHMDSDSMICAGNLQKPG-----QPTCOG 507  
 DB 329 HEALSPFYSERLKEAHVRLYPSRCTSQHLLNRTVTDNMLCAGDTRSGGPGQANLHDACOG 388  
 QY 508 DSGGPLTCEKDGTYVYVGVISWGLECGKR--PGVYTVQTKFLNWKATIK 555  
 DB 389 DSGGPLVCLDGRMTLVGIISWGLGCGQKDVPGVYTKVTNYLDWIRDNNMR 438

RESULT 10  
 AAR13922  
 ID AAR13922 standard; protein; 434 AA.  
 XX  
 AC AAR13922;  
 XX  
 DT 09-JAN-2003 (revised)  
 DT 27-NOV-1991 (first entry)  
 XX  
 DE Delta(92-179) and delta(466-470) tPA variant.  
 XX  
 KW tissue plasminogen activator; infarction; coagulation.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 1..39  
 FT /label= Finger\_domain  
 FT Domain 92..173  
 FT /label= Kringle\_2  
 FT  
 XX WO9113149-A.  
 XX  
 XX 05-SEP-1991.  
 XX  
 XX 01-MAR-1990; 90US-00486657.  
 XX

PR 01-MAR-1990; 90US-00486657.  
XX (GETH ) GENENTECH INC.  
XX Gill JF, Presta LG, Zoller ML;  
XX WPI; 1991-281468/38.  
DR  
XX New tissue plasminogen activator variant used for vascular conditions -  
PT e.g. to prevent fibrin deposition of adhesion formation of reformation,  
PT deep vein thrombosis, peripheral vascular disease, embolism etc.  
XX  
XX Claim 12; Page 27; 33pp; English.  
XX  
XX This tPA variant has a deletion of amino acids Pro(466) to Leu(470) of  
CC the corresponding wild-type. The deletion is in the serine protease  
CC domain, making the variant more fibrin specific than the wild-type tPA.  
CC The variant also has a deletion of amino acids Cys(92) to Asp(179) of the  
CC wild-type sequence (i.e. Kingle 1 domain). See also AAR13917-R13921 and  
CC AAR14486. (Updated on 09-JAN-2003 to add missing OS field.)  
XX  
XX Sequence 434 AA;  
SQ  
Query Match 25.3%; Score 798.5; DB 2; Length 434;  
Best Local Similarity 37.6%; Pred. No. 5.4e-40;  
Matches 175; Conservative 70; Mismatches 165; Indels 55; Gaps 14;  
QY 112 QNTCKDNPCGRCGLITQSPPPYR-----CVCKHPYTGSCSOWPV--CRENPC 159  
DB 3 QVICRDE---KTQMIYQCHOSWLRPVLRSNRVEYWCN---SGRAQCHSVPVKSCSEPRC 56  
QY 160 QNGATCSRHRKSKFTACPDQFKGKFCFEGS-DDCYVGDGYSYRGKNNRTVNOHACLW 218  
DB 57 FNGGTCQALYFSDFCVQCPEGAGKCEIDTRATCYFGNGSAVYRGTHSLTESGASCLPW 116  
QY 219 NSHLLLOENYMFEMDAETHGIGENFCRNPDADEKPCFKIKVNDKVKWEYCDVSACA 278  
DB 117 NSMLLIGKYTAQNSAOLGLGKKNYCNPDGDAKPNCHV-LKMRRLTWEYCDVPSCS- 174  
QY 279 QDVAYPEESFTPESTKLPQFDCSGCTEATBKIRIYCGFKSTAGKIPQWASLSLPLT 338  
DB 175 -----TCGLRQXSQPOF-RIKGGLFADIASHFPWQAAIFAK--HR 210  
QY 339 ISMPOGHFCGALIHPCWVLTAAHCTD--IKTRHLKVLVGLDQDLKEEFHQSPRVEKIF 396  
DB 211 RSPGERFLCGGILISSCWLSAAHCTQERFPFPHLTVILGRYRVVPGEEQKEVEKYI 270  
QY 397 KYSHYNERDEIPHNDIALKLPVDGHCALBSKYVTKVCLP--DGSFPGSGECHISGWV 454  
DB 271 VKXEPD--DDTYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTCELSGYK 328  
QY 455 TETGKG--SRQLIDAKVKLIANTLNSRQLYDHMDSDMICAGNLQKPGQDTCCQDGGP 512  
DB 329 HEALSPFYSERLKEAHVRLYPSRCTSQHLLNRTVTDNMLCAGDTRSGHDAQCGDGGP 388  
QY 513 LTCCKDGTYYVGVISWGLECGKR--PGVYTOVTKFLMWIKATK 555  
DB 389 LVCNDGRMTLVGIISWGLGCGQKRDVPGVYTKVNTYLDWIRDNR 433  
RESULT 11  
AAR09261  
ID AAR09261 standard; protein; 439 AA.  
XX  
AC AAR09261;  
XX  
XX 24-FEB-1993 (first entry)  
DT  
XX t-PA variant d92-179, I210R, G211A, K212R, V213R, T252R, F305H.  
DE  
XX Tissue plasminogen activator; zymogen; clot; plasma; plasmin.  
KW  
XX Homo sapiens.  
OS

XX Key Location/Qualifiers  
FH Misc-difference 91..92  
FT /note= "amino acids 92-179 of the wild-type, mature t-PA  
FT have been deleted"  
FT Misc-difference 122  
FT /label= substitution  
FT /note= "I210 of the wild-type, mature t-PA has been  
FT substituted for R"  
FT Misc-difference 123  
FT /label= substitution  
FT /note= "G211 of the wild-type, mature t-PA has been  
FT substituted for A"  
FT Misc-difference 124  
FT /label= substitution  
FT /note= "K212 of the wild-type, mature t-PA has been  
FT substituted for R"  
FT Misc-difference 125  
FT /label= substitution  
FT /note= "V213 of the wild-type, mature t-PA has been  
FT substituted for R"  
FT Misc-difference 164  
FT /label= substitution  
FT /note= "T252 of the wild-type, mature t-PA has been  
FT substituted for R"  
FT Misc-difference 217  
FT /label= substitution  
FT /note= "F305 of the wild-type, mature t-PA has been  
FT substituted for H"  
XX  
XX WO9002798-A.  
PN  
XX 22-MAR-1990.  
PD  
XX 02-SEP-1988; 88US-00240856.  
PF  
XX 02-SEP-1988; 88US-00240856.  
PR  
XX 24-JUL-1989; 89US-00384608.  
PR  
XX (GETH ) GENENTECH INC.  
PA  
XX  
XX Anderson S, Bennett WE, Botstein D, Higgins DL, Paoni NF;  
PI Zoller M;  
XX  
XX WPI; 1990-115987/15.  
DR  
XX Modified tissue plasminogen activator - activated only when proximate to  
PT plasmin at site of clot and not systemically.  
XX  
XX Claim 17; Page 44; 63pp; English.  
XX  
XX The t-PA variants represented in AAR09215-63 (and the T252R or N184S  
CC analogues thereof or combinations thereof) and AAR09266-83 have their  
CC fibrinolytic activity turned off when generally in the plasma and  
CC activated when proximate to plasmin at the site of the clot thus  
CC providing specific localised clot therapy  
XX  
XX Sequence 439 AA;  
SQ  
Query Match 25.3%; Score 797; DB 2; Length 439;  
Best Local Similarity 37.4%; Pred. No. 6.7e-40;  
Matches 176; Conservative 69; Mismatches 165; Indels 60; Gaps 15;  
QY 112 QNTCKDNPCGRCGLITQSPPPYR-----CVCKHPYTGSCSOWPV--CRENPC 159  
DB 3 QVICRDE---KTQMIYQCHOSWLRPVLRSNRVEYWCN---SGRAQCHSVPVKSCSEPRC 56  
QY 160 QNGATCSRHRKSKFTACPDQFKGKFCFEGS-DDCYVGDGYSYRGKNNRTVNOHACLW 218  
DB 57 FNGGTCQALYFSDFCVQCPEGAGKCEIDTRATCYFGNGSAVYRGTHSLTESGASCLPW 116  
QY 219 NSHLLLOENYMFEMDAETHGIGENFCRNPDADEKPCFKIKVNDKVKWEYCDVSACA 278  
DB  
OS

Db 117 NSMILPAREYTAQNPSAQAALGLGKKNYCNPNPDGDAKPWCHV-LKNRRLRWECYDVPSCS- 174

QY 279 QDVAYPEESPTSTKLPFDSCGKTEIAERKIKYIGFKSTAGKHPWQASLOSSLPLT 338

Db 175 -----TCGLRQYSQPF-RKGGFLFADIASHPWQAIFAK--HR 210

QY 339 ISMPOGHFCGGLIHPCWLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSPRVEKIF 396

Db 211 RSPGERHLCGGLISSCWILSAHCFQERPPPHLTVILGRYRVVPGESEQFEVEKYI 270

QY 397 KYSHYNERDEIPHNDIALKLPVDGHCALESKYKVTCLP--DGSFPGSGECHISGMV 454

Db 271 VHKEFD--DDTYDNDIALQLKSDSRCAQESSVVRTVCLPADLQLPDWTECELSGYGK 328

QY 455 TETGKG--SRQLLDKVKLIANTLNSROLYDHMDDSMICAGNLQKPG-----QDTCOG 507

Db 329 HEALSPFYSERLKEAHVRYPSCRTSQHLLNRTVDMCLAGDTRSGGPQANLHDACQ 388

QY 508 DSGGPLTCEXDTGVYVYGVISWGLECGKR--PGVYTQVTKFLNWKATIK 555

Db 389 DSGGPLVCLNDGRMTLVGIISWGLGCGQKDPGVYTKVTNYLDWIRDNR 438

RESULT 12

AAR09258

ID AAR09258 standard; protein; 439 AA.

AC AAR09258;

DT 24-FEB-1993 (first entry)

XX t-PA variant d92-179, F305H.

XX Tissue plasminogen activator; zymogen; clot; plasma; plasmin.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 91..92

FT /note= "amino acids 92-179 of the wild-type, mature t-PA

FT have been deleted"

FT Misc-difference 217

FT /label= substitution

FT /note= "F305 of the wild-type, mature t-PA has been

FT substituted for H"

XX WO9002798-A.

XX 22-MAR-1990.

XX 02-SEP-1988; 88US-00240856.

XX 02-SEP-1988; 88US-00240856.

XX 24-JUL-1989; 89US-00384608.

XX (GETH ) GENENTECH INC.

XX Anderson S, Bennett WE, Botstein D, Higgins DL, Paoni NF;

XX Zoller M;

XX WPI; 1990-115987/15.

XX Modified tissue plasminogen activator - activated only when proximate to

XX plasmin at site of clot and not systemically.

XX Claim 17; Page 44; 63pp; English.

XX The t-PA variants represented in AAR09215-63 (and the T252R or N184S

XX analogues thereof or combinations thereof) and AAR09266-83 have their

XX fibrinolytic activity turned off when generally in the plasma and

XX activated when proximate to plasmin at the site of the clot thus

XX providing specific localised clot therapy

SQ Sequence 439 AA;

Query Match 25.2%; Score 795; DB 2; Length 439;

Best local Similarity 37.4%; Pred. NO. 8.8e-40;

Matches 176; Conservative 70; Mismatches 164; Indels 60; Gaps 15;

QY 112 QNTCKNPNCRGQCLITQSPPPYR-----CVCKHPYTGSPSCQVVPV--CRPNPC 159

Db 3 QVICRDE--KTQWVQQHQSGLRPVLRNSRVYCNW---SGRAQCHSVPVKSCBPRC 56

QY 160 QNGATSRHRKRSKFTACPDQFKGFECEIGS--DDCYVGDGYSGYRGMNRTVNHACLYW 218

Db 57 FNGGTCQQAALYFSDVFCQCEGPFAGKCEIDTRATCYFNGNSAYRGTSHLTESGASCLPW 116

QY 219 NSHLLLOENNMFMEDAEHTGIGEHNFRCNPFDADEKPCWCFIKVTNDKVKWEYCDVSACSA 278

Db 117 NSMILIKVYTAQNPASAQALGLGKKNYCNPNPDGDAKPWCHV-LKNRRLTWECYDVPSCS- 174

QY 279 QDVAYPEESPTSTKLPFDSCGKTEIAERKIKYIGFKSTAGKHPWQASLOSSLPLT 338

Db 175 -----TCGLRQYSQPF-RKGGFLFADIASHPWQAIFAK--HR 210

QY 339 ISMPOGHFCGGLIHPCWLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSPRVEKIF 396

Db 211 RSPGERHLCGGLISSCWILSAHCFQERPPPHLTVILGRYRVVPGESEQFEVEKYI 270

QY 397 KYSHYNERDEIPHNDIALKLPVDGHCALESKYKVTCLP--DGSFPGSGECHISGMV 454

Db 271 VHKEFD--DDTYDNDIALQLKSDSRCAQESSVVRTVCLPADLQLPDWTECELSGYGK 328

QY 455 TETGKG--SRQLLDKVKLIANTLNSROLYDHMDDSMICAGNLQKPG-----QDTCOG 507

Db 329 HEALSPFYSERLKEAHVRYPSCRTSQHLLNRTVDMCLAGDTRSGGPQANLHDACQ 388

QY 508 DSGGPLTCEXDTGVYVYGVISWGLECGKR--PGVYTQVTKFLNWKATIK 555

Db 389 DSGGPLVCLNDGRMTLVGIISWGLGCGQKDPGVYTKVTNYLDWIRDNR 438

RESULT 13

AAR09259

ID AAR09259 standard; protein; 439 AA.

AC AAR09259;

DT 24-FEB-1993 (first entry)

XX t-PA variant d92-179, I210R, G211A, K212R, V213R, F305H.

XX Tissue plasminogen activator; zymogen; clot; plasma; plasmin.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 91..92

FT /note= "amino acids 92-179 of the wild-type, mature t-PA

FT have been deleted"

FT Misc-difference 122

FT /label= substitution

FT /note= "I210 of the wild-type, mature t-PA has been

FT substituted for R"

FT Misc-difference 123

FT /label= substitution

FT /note= "G211 of the wild-type, mature t-PA has been

FT substituted for A"

FT Misc-difference 124

FT /label= substitution

FT /note= "K212 of the wild-type, mature t-PA has been

FT substituted for R"

FT Misc-difference 125

FT /label= substitution

FT /note= "V213 of the wild-type, mature t-PA has been

FT substituted for R"

ID	XX	AAP71742 standard; protein; 623 AA.	
XX	XX		
AC	AC	AAP71742;	
XX	XX		
DT	DT	07-JUN-1991 (first entry)	
XX	XX		
DE	DE	Ile(277)t-PA with extra finger- and extra growth factor-domains.	
XX	XX		
KW	KW	tissue plasminogen activator; finger domain; fibrinolysis.	
XX	XX		
OS	OS	Synthetic.	
XX	XX		
FH	FH	Location/Qualifiers	
FT	FT	9..46	
FT	FT	/label= extra finger domain	
FT	FT	54..87	
FT	FT	/label= extra growth factor domain	
FT	FT	97..623	
FT	FT	/label= natural human tPA	
FT	FT	/note= "Lys(277) replaced by Ile"	
XX	XX		
PN	PN	EP241210-A.	
XX	XX		
PD	PD	14-OCT-1987.	
XX	XX		
PF	PF	01-APR-1987; 87EP-00302837.	
XX	XX		
PR	PR	02-APR-1986; 86GB-00008020.	
XX	XX		
PA	PA	(BEEC) BEECHAM GROUP PLC.	
XX	XX		
PI	PI	Browne MJ, Robinson JH;	
XX	XX		
DR	DR	WPI; 1987-286054/41.	
XX	XX		
PT	PT	Modified fibrinolytically active plasminogen activator - has added finger	
PT	PT	domain(s) for antithrombotic use and is obtd. by recombinant DNA	
PT	PT	procedures.	
XX	XX		
PS	PS	Claim 11; Page 12; 16pp; English.	
XX	XX		
CC	CC	The modified tPA molecules retain the fibrinolytic activity of natural	
CC	CC	tPA and are used to treat thrombotic diseases. The extra finger domain	
CC	CC	extends at the N-terminus to comprise amino acids 1-5 of native tPA. The	
CC	CC	tPA mutcin can form part of a hybrid protein when linked to another	
CC	CC	fibrinolytically active protease so that the catalytic site of the hybrid	
CC	CC	protein essential for fibrinolytic activity is optionally blocked by a	
XX	XX	removable blocking group. See also AAP71740-1 and AAP71743-5	
XX	XX		
SQ	SQ	Sequence 623 AA;	
		Query Match 25.1%; Score 793; DB 1; Length 623;	
		Best Local Similarity 31.2%; Pred. No. 1.6e-39;	
		Matches 188; Conservative 75; Mismatches 182; Indels 158; Gaps 17	
Qy	77	CQNPCEHGGDC--LVHGSTFTCSCLAFFSGNKCQ-----KVQNTCKDNPCGRGQCL 126	
Db	54	CSEPRCFNGGTCCQALYFDFVCCQPSGFAGKCEIDTRATPGSYQVCRDE--KTQMI 110	
Qy	127	ITQSPPYR-----CVCKHPTTPGSCSQWVY--CRNPFCQNGATCSRHKRSKF 174	
Db	111	YQGHOSWLRPVLRNSRVEYWCN---SGRAQCHSVFVKCSBPRCFNGGTCCQALYFDF 167	
Qy	175	TCACPDQKGFCEI-----	
Db	168	VQCEGEGAGKCEIDTRATCYEDQGISYRGTWSTAEGACTWNSSALAAQKPYSGRRP 227	
Qy	190	-----GSDDCYVGDGYSGK 205	
Db	228	DAIRLGLGNHNYCRNPDRSKWCYVPKAGKYSFECSTPACSEGNDCYFGNGSAYRGT 287	
Qy	206	MRTVYNQACLYNWSHLLQLQNYNMFMEDAETHGIGEHNFRCRNPDADEKFWCFIKVTNDK 265	



Db 288 HSLTESGASCLPWNMILLIGKYYTAQNPASQAALGLGKHNVCENPDGDAKPMCHV-LKNRR 346  
Qy 266 VKWEYCDVSACSADQVAYPEESPTFSTKLPDFDSCGKTEIAERKIKRIYGGFKSTAGKH 325  
Db 347 LTWEYCDVPSCS-----TCGLRQYSQPQF-RIIGGLFADIASH 383  
Qy 326 PQWASLQSLPLTISMPQGHFCGALIHPCWVLTAAHCTD--IKTRHLKVLGDDQLKKE 383  
Db 384 PQWAAIFAK--HRRSPGERFLCGGILISSCWLLSAHKFOERPPPHLTVILGRYRVVP 441  
Qy 384 EFHEQSFVEKIPKYSHYNERDEIPHNDIALKLPVDGHCALSKYKVTVCPL--DGSF 441  
Db 442 GEEOQKFEVEKIVHKEFD--DDTYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQ 499  
Qy 442 PSGSECHISGWVTETGKG--SRQLLDKAVKLIANTLNSRQLYDHMIDDSMICAGNLQK 499  
Db 500 PDWTECELSGKGHEALSFFYERLKEAHRVLYPSRCSCTSHLLNRTVTDNMLCAGDTRS 559  
Qy 500 PG-----QDTCCGDSGGPLTCEKDGTYVYVGVISWGLECGKR--PGVYTVTKFLNWIKA 552  
Db 560 GGPOANLHDACQDGGGGLVCLNDGRMTLVGIISWGLGCGQKDVPGVYTKVTNYLDWIRD 619  
Qy 553 TIK 555  
Db 620 NMR 622

RESULT 15  
AAP71740  
ID AAP71740 standard; protein; 623 AA.  
AC AAP71740;  
XX 07-JUN-1991 (first entry)  
DT t-PA with extra finger domain and extra growth factor domain.  
DE tissue plasminogen activator; finger domain; fibrinolysis.  
XX Synthetic.  
OS Key Location/Qualifiers  
FH Domain 9..46  
FT Domain /label= extra finger domain  
FT Domain 54..87  
FT Protein /label= extra growth factor domain  
FT 97..623  
FT /label= natural human tPA  
PN EP241210-A.  
XX 14-OCT-1987.  
XX 01-APR-1987; 87EP-00302837.  
XX 02-APR-1986; 86GB-00008020.  
XX (BEEC ) BEECHAM GROUP PLC.  
XX Browne MJ, Robinson JH;  
XX WPI; 1987-286054/41.  
XX Modified fibrinolytically active plasminogen activator - has added finger  
PT domain(s) for antithrombotic use and is obtd. by recombinant DNA  
PT procedures.  
XX Claim 11; Page 12; 16pp; English.

XX The modified tPA molecules retain the fibrinolytic activity of natural  
CC tPA and are used to treat thrombotic diseases. The extra finger domain  
CC extends at the N-terminus to comprise amino acids 1-5 of native tPA. The  
CC tPA mutein can form part of a hybrid protein when linked to another

CC fibrinolytically active protease so that the catalytic site of the hybrid  
CC protein essential for fibrinolytic activity is optionally blocked by a  
XX removable blocking group. See also AAP71741-5  
SQ Sequence 623 AA;  
Query Match 25.1%; Score 792; DB 1; Length 623;  
Best Local Similarity 31.2%; Pred. No. 1.9e-39;  
Matches 188; Conservative 75; Mismatches 182; Indels 158; Gaps 17;  
Qy 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSNGKCO-----KVONTCKNDKPCGGQCL 126  
Db 54 CSBPCFNGGTCCQALYFSDFCQCPGEGFAGKCCETDTRATPGSYQVICRDE---RTQMI 110  
Qy 127 ITQSPPYR-----CVCKHPYTPGSCSQVVPV--CRPNPCONGATCSRHKRSKF 174  
Db 111 YQHQSWLRVLRNSRVEYWCN---SGRAQCHSVFKSCSEPCRCFNGGTCCQALYFSD 167  
Qy 175 TCACPDQFKKFCBI----- 189  
Db 168 VCQCPGEGFAGKCCETDTRATCYEDQGISYRGWTSTAESGAECTNMNSSLAAQKPYSGRRP 227  
Qy 190 -----GSDDCYVGDGYSYRGK 205  
Db 228 DAIRLGLGNHNYCRNDRDSKPMCYVFKAGYSEFCSTPACSEGNSDCYFGNGSAVRGT 287  
Qy 206 MNRTVNOHACLYWNSHLLLOENYMFEMDAETHIGIEHNFCRNPDADEKPCFIKVTNDK 265  
Db 288 HSLTESGASCLPWNMILLIGKYYTAQNPASQAALGLGKHNVCENPDGDAKPMCHV-LKNRR 346  
Qy 266 VKWEYCDVSACSADQVAYPEESPTFSTKLPDFDSCGKTEIAERKIKRIYGGFKSTAGKH 325  
Db 347 LTWEYCDVPSCS-----TCGLRQYSQPQF-RINGGLFADIASH 383  
Qy 326 PQWASLQSLPLTISMPQGHFCGALIHPCWVLTAAHCTD--IKTRHLKVLGDDQLKKE 383  
Db 384 PQWAAIFAK--HRRSPGERFLCGGILISSCWLLSAHKFOERPPPHLTVILGRYRVVP 441  
Qy 384 EFHEQSFVEKIPKYSHYNERDEIPHNDIALKLPVDGHCALSKYKVTVCPL--DGSF 441  
Db 442 GEEOQKFEVEKIVHKEFD--DDTYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQ 499  
Qy 442 PSGSECHISGWVTETGKG--SRQLLDKAVKLIANTLNSRQLYDHMIDDSMICAGNLQK 499  
Db 500 PDWTECELSGKGHEALSFFYERLKEAHRVLYPSRCSCTSHLLNRTVTDNMLCAGDTRS 559  
Qy 500 PG-----QDTCCGDSGGPLTCEKDGTYVYVGVISWGLECGKR--PGVYTVTKFLNWIKA 552  
Db 560 GGPOANLHDACQDGGGGLVCLNDGRMTLVGIISWGLGCGQKDVPGVYTKVTNYLDWIRD 619  
Qy 553 TIK 555  
Db 620 NMR 622

Search completed: May 24, 2004, 09:46:50  
Job time : 56 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 09:43:55 ; Search time 16.5 Seconds  
(without alignments)  
3264.682 Million cell updates/sec

Title: US-09-912-559-3

Perfect score: 3154

Sequence: 1 MFARMSDLHLLMALVGKT.....TQVTKFLNWKATIKSBSGF 560

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3154	100.0	560	1 JC4795	plasma hyaluronan-
2	2528	80.2	558	2 JC5978	plasma hyaluronan-
3	832.5	26.4	555	1 A46888	hepatocyte growth
4	773.5	24.5	603	2 S28941	coagulation factor
5	766.5	24.3	562	1 UKHUT	t-plasminogen acti
6	757.5	24.0	559	1 A35029	t-plasminogen acti
7	745.5	23.6	559	1 A29941	t-plasminogen acti
8	741	23.5	615	1 KFHU12	coagulation factor
9	731	23.2	477	1 A34369	t-plasminogen acti
10	731	23.2	477	2 JS0598	t-plasminogen acti
11	720.5	22.8	431	2 JS0599	t-plasminogen acti
12	706.5	22.4	593	2 S45281	coagulation factor
13	704.5	22.3	477	2 JS0597	t-plasminogen acti
14	704	22.3	432	1 S18932	u-plasminogen acti
15	686	21.8	442	1 UKPG	u-plasminogen acti
16	683.5	21.7	433	1 JN0560	u-plasminogen acti
17	680	21.6	433	1 UKMS	u-plasminogen acti
18	660	20.9	433	1 UKBAY	u-plasminogen acti
19	660	20.9	812	1 PLBO	plasmin (EC 3.4.21
20	658.5	20.9	431	1 UKHU	u-plasminogen acti
21	650.5	20.6	434	1 A35005	u-plasminogen acti
22	643.5	20.4	394	2 JS0600	u-plasminogen acti
23	642.5	20.4	810	1 PLHU	plasmin (EC 3.4.21
24	635.5	20.1	810	2 B30849	plasmin (EC 3.4.21
25	607	19.2	790	1 PLPG	plasmin (EC 3.4.21
26	598.5	19.0	812	1 PLMS	apoprotein(a) (EC
27	592.5	18.8	4548	1 SC0657	apoprotein(a) (EC
28	587	18.6	455	2 A61545	plasmin (EC 3.4.21
29	578.5	18.3	1420	2 A32869	apolipoprotein(a)

30	571.5	18.1	810	2 I46260	plasmin (EC 3.4.21
31	557	17.7	460	2 B61545	plasmin (EC 3.4.21
32	467.5	14.8	475	1 EXCH	coagulation factor
33	465	14.7	716	1 JC5061	macrophage-stimula
34	462.5	14.7	761	2 JC5759	brain-specific ser
35	461	14.6	716	1 A40332	macrophage-stimula
36	457.5	14.5	728	1 A60185	hepatocyte growth
37	456.5	14.5	728	1 A35644	hepatocyte growth
38	444	14.1	710	1 IS1283	coagulation factor
39	439.5	13.9	461	1 KFHU	macrophage-stimula
40	438.5	13.9	711	1 A47136	membrane-bound arg
41	438.5	13.9	855	2 JC7731	thrombin (EC 3.4.2
42	437.5	13.9	625	1 TBBO	thrombin (EC 3.4.2
43	429	13.6	622	1 TBHU	coagulation factor
44	427	13.5	482	1 EXRT	hepatocyte growth
45	423.5	13.4	728	1 JH0579	hepatocyte growth

## ALIGNMENTS

### RESULT 1

JC4795

plasma hyaluronan-binding protein precursor - human

N:Alternate names: hepatocyte growth factor activator-like protein; PHBP

N:Contains: serine proteinase (EC 3.4.21.-)

C:Species: Homo sapiens (man)

C>Date: 15-Oct-1995 #sequence\_revision 16-Aug-1996 #text\_change 19-Jul-2002

C:Accession: JC4795

J:Choi-Miura, N.H.; Tobe, T.; Sumiya, J.; Nakano, Y.; Sano, Y.; Mazda, T.; Tomita, M.

J:Biochem. 119, 1157-1165, 1996

A:Title: Purification and characterization of a novel hyaluronan-binding protein (PHBP)

A:Reference number: JC4795; MUID:96425001; PMID:8827452

A:Accession: JC4795

A:Molecule type: mRNA

A:Residues: 1-560 <CRO>

A:Cross-references: NID:G1836158; PIDN:AAB46909.1; PID:G1836159

A:Experimental source: plasma

A:Note: parts of this sequence, including the amino ends of the mature chains, were dete

C:Genetics:

A:Gene: GDB:HABP2; HABP; PHBP; HGFAL

A:Cross-references: GDB:4573962

C:Complex: a disulfide-bonded heterodimer of chains produced from the same precursor; th

C:Superfamily: plasma hyaluronan-binding protein; EGF homology; kringle homology; trypsi

C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; hyaluronic acid; hydrolase;

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-313/Product: plasma hyaluronan-binding protein, 50K chain #status predicted <50K>

F:77-108/Domain: EGF homology <EG1>

F:115-147/Domain: EGF homology <EG2>

F:194-276/Domain: EGF homology <EG3>

F:314-516/Domain: trypsin homology <TRY>

F:54,207/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:77-88,82-97,99-108,115-125,120-136,138-147,154-165,159-176,178-187,194-276,215-257,246

F:362,405,509/Active site: His, Asp, Ser #status predicted

Query Match 100.0%; Score 3154; DB 1; Length 560;

Best Local Similarity 100.0%; Pred. No. 3e-188;

Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHLLMALVGKTACGFSLSLLESIDPDWTPDQYDYSYEDYNOENTSSLT 60

DB 1 MFARMSDLHLLMALVGKTACGFSLSLLESIDPDWTPDQYDYSYEDYNOENTSSLT 60

QY 61 HAENPDWYTEDQADPCQPNPCHEGGDCLVHGSTFTCSCLAPFSGNKCQKVQNTCKDNPC 120

DB 61 HAENPDWYTEDQADPCQPNPCHEGGDCLVHGSTFTCSCLAPFSGNKCQKVQNTCKDNPC 120

QY 121 GRGCLITQSPPYRVCVKHPYTGPSCSQVVPVCRPNPCQNGATCSRHKRSRFTACPD 180

DB 121 GRGCLITQSPPYRVCVKHPYTGPSCSQVVPVCRPNPCQNGATCSRHKRSRFTACPD 180

QY 181 QPKGFCFCEISGDDCYVGDGYSGYRGKQNRVTNQHACLYWNHSHLLQENYNNMFWEAETHGI 240  
DB 181 QPKGFCFCEISGDDCYVGDGYSGYRGKQNRVTNQHACLYWNHSHLLQENYNNMFWEAETHGI 240  
QY 241 GHNFCRNPDADEKPCFETKVTNDKVKWEYCDVSACSADVAYPESTPSTKLPFGDS 300  
DB 241 GHNFCRNPDADEKPCFETKVTNDKVKWEYCDVSACSADVAYPESTPSTKLPFGDS 300  
QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSSLPLTISMPQGHFCGCGALHPCWVLT 360  
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSSLPLTISMPQGHFCGCGALHPCWVLT 360  
QY 361 AHCTDIKTRHLKVVLDGQDLKKEEFHEQSFVEKIFKYSYNERDEI PHNDIALALKPV 420  
DB 361 AHCTDIKTRHLKVVLDGQDLKKEEFHEQSFVEKIFKYSYNERDEI PHNDIALALKPV 420  
QY 421 DGHCALESKYKVTCLPDGSPGSECHISGMWVETGKSRQLDAKVKLIANTLNSR 480  
DB 421 DGHCALESKYKVTCLPDGSPGSECHISGMWVETGKSRQLDAKVKLIANTLNSR 480  
QY 481 QLYDHMLDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVGVSWGLECGKRPVY 540  
DB 481 QLYDHMLDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVGVSWGLECGKRPVY 540  
QY 541 TQVTKFLNWKATIKSES 560  
DB 541 TQVTKFLNWKATIKSES 560

RESULT 2  
JC5878  
C:Species: Mus musculus (house mouse)  
C:Date: 11-Mar-1998 #sequence\_revision 11-Mar-1998 #text\_change 16-Jul-1999  
C:Accession: JC5878  
R:Hashimoto, K.; Tobe, T.; Sumiya, J.; Saguchi, K.; Sano, Y.; Nakano, Y.; Choi-Miura, N.  
Biochem. Biophys. Res. Commun. 20, 1127-1130, 1997  
A:Title: Cloning of the cDNA for a mouse homologue of human PHBP: A novel hyaluronan-bin  
A:Reference number: JC5878; MUID:98065239; PMID:9401717  
A:Accession: JC5878  
A:Molecule type: mRNA  
A:Residues: 1-558 <HAS>  
C:Comment: This protein acts as serine protease.  
C:Superfamily: Plasma hyaluronan-binding protein; EGF homology; kringle homology; trypsin  
F:1-23/Domain: Signal sequence #status predicted <SIG>  
F:24-311/Product: Plasma hyaluronan-binding protein large chain #status predicted <MATL>  
F:75-106/Domain: EGF homology <EG1>  
F:113-145/Domain: EGF homology <EG2>  
F:152-186/Domain: EGF homology <EG3>  
F:192-274/Domain: kringle homology <KE1>  
F:312-558/Product: Plasma hyaluronan-binding protein small chain #status predicted <MATS>  
F:312-548/Domain: trypsin homology <TRY>

Query Match 80.2%; Score 2528; DB 2; Length 558;  
Best Local Similarity 78.5%; Pred. No. 1.8e-149;  
Matches 439; Conservative 39; Mismatches 79; Indels 2; Gaps 1;  
QY 1 MFARMSDLHVLALLMALVCKTACGSLMSLESLLPDPDWTPODYDYSDYDYNQENTSTLT 60  
DB 1 MFVRLVRLVRLVLLALVGVKSVIGSLMSFIAPPDWTPTDDYIYYSQSFDEDPSTQT 60  
QY 61 HAENPDWYTTEDQADPCQPNCEHGGDCLVHGSTFTCSCLAPFSGNCKQVQNTCKDNPC 120  
DB 61 TPENPDWYTTED--DPCQSNCEHGGDCLIRGDTFSCSPAPFSGSRQCAQNKCKDNPC 118  
QY 121 GRGQLITQSPPYRCVKHPTGPGSCQVVPVCRPNCPNGATCSRKRKSTACPD 180  
DB 119 VHGDCLITQKHPIYRCACKPYTGPGSCQVVPVCRPNCPNGATCSRKRKSTACPD 178  
QY 181 QPKGFCFCEISGDDCYVGDGYSGYRGKQNRVTNQHACLYWNHSHLLQENYNNMFWEAETHGI 240  
DB 179 QYKGFCEIGPDDCYVGDGYSGYRGKQNRVTNQHACLYWNHSHLLQENYNNMFWEAETHGI 238

QY 241 GEHNFCRNPDADEKPCFETKVTNDKVKWEYCDVSACSADVAYPESTPSTKLPFGDS 300  
DB 239 AEHNFCRNPDADEKPCFETKVTNDKVKWEYCDVTVCVDPDTPNPVESLLEPYWELPFGFS 298  
QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSSLPLTISMPQGHFCGCGALHPCWVLT 360  
DB 299 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSSLPLTISMPQGHFCGCGALHPCWVLT 358  
QY 361 AHCTDIKTRHLKVVLDGQDLKKEEFHEQSFVEKIFKYSYNERDEI PHNDIALALKPV 420  
DB 359 AHCTDIKTRHLKVVLDGQDLKKEEFHEQSFVEKIFKYSYNERDEI PHNDIALALKPV 418  
QY 421 DGHCALESKYKVTCLPDGSPGSECHISGMWVETGKSRQLDAKVKLIANTLNSR 480  
DB 419 GGHCALESKYKVTCLPDGSPGSECHISGMWVETGKSRQLDAKVKLIANTLNSR 478  
QY 481 QLYDHMLDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVGVSWGLECGKRPVY 540  
DB 479 QLYDHTDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVGVSWGLECGKRPVY 538  
QY 541 TQVTKFLNWKATIKSES 559  
DB 539 TQVTKFLNWKATIKSES 557

RESULT 3  
A46688  
C:Species: Homo sapiens (man)  
C:Date: 21-Sep-1993 #sequence\_revision 25-Aug-1995 #text\_change 08-Dec-2000  
C:Accession: A46688  
R:Yamazawa, K.; Shimomura, T.; Kitamura, A.; Kondo, J.; Morimoto, Y.; Kitamura, N.  
J. Biol. Chem. 268, 10024-10028, 1993  
A:Title: Molecular cloning and sequence analysis of the cDNA for a human serine proteas  
d coagulation factor XII.  
A:Reference number: A46688; MUID:93252878; PMID:7693665  
A:Accession: A46688  
A:Molecule type: mRNA  
A:Residues: 1-855 <MIY>  
A:Cross-references: DDBJ:D14012; NID:9219680; PIDN:BA003113.1; PID:9219681  
A:Experimental source: liver (mRNA); serum (protein)  
A:Note: sequence extracted from NCBI backbone (NCBIN:131227, NCBI:131228)  
A:Note: parts of the sequence, including the amino ends of the heavy and light chains,  
C:Genetics:  
A:Gene: GDB:HGPA; HGPA: HGPA  
A:Cross-references: GDB:9954514  
A:Map position: 4p16-4p16  
C:Function:  
A:Description: activates hepatocyte growth factor by specific proteolytic cleavage  
A:Pathway: tissue repair and regeneration  
C:Superfamily: coagulation factor XII; EGF homology; fibronectin type I repeat homology  
C:Keywords: glycoprotein; hydrolase; kringle; liver; plasma; serine proteinase  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:108-148/Domain: fibronectin type II repeat homology <IF2>  
F:164-197/Domain: EGF homology <EG1>  
F:202-237/Domain: fibronectin type I repeat homology <IF1>  
F:245-278/Domain: EGF homology <EG2>  
F:286-367/Domain: kringle homology <KR>  
F:373-407/Product: hepatocyte growth factor activator light chain #status experimental  
F:408-655/Product: hepatocyte growth factor activator heavy chain #status experimental  
F:408-641/Domain: trypsin homology <TRY>  
F:408-489,489,492,546/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:164-175,169-186,188-197,202-230,228-237,245-256,250-267,269-278,286-367,307-349,338-3  
F:447-497,598/Active site: His, Asp, Ser #status predicted

Query Match 26.4%; Score 832.5; DB 1; Length 655;  
Best Local Similarity 33.2%; Pred. No. 2.5e-44;  
Matches 192; Conservative 87; Mismatches 216; Indels 83; Gaps 21;  
QY 31 ESUDPWTPODYDYSDYDYNQENTSTLTAEHNDP-----YTTEDQA----- 74  
DB 99 QALTEDGRCPFRFFYGGRLHACTSEGAHRK---WCATTHYDRDRAGVCVEATPP 155

75 QY -----DFCPNPFCEHGDC--LVHGSTFTCSCLAPFSGNKKCKQVQNTCKNDKPNCGRQCL 126  
156 Db GPPAALDFCASPCLNGSCSNTQDPOSHYSCSPRAFTG-----KD--CGTEKCF 203  
127 QY ITQSPYY-----RCVCKHPVTGSCSQV--PVCRRNPONGATCSRHK 169  
204 Db DETREYLEGDRWARVRQHVQCEC---FGRTWCEGTRHTRACLSPLCLNGTCHLIV 260  
170 QY RRSKFTACPDQFGKFCFCEIGSD- CYVGDGYSYRGKRNRTVNOHACLYNWSHLLQENY 228  
261 Db ATGTTVCACPPGFAGRLCNIEPDERCFGLNGTGYRGVASTSASGLSCLAWSNLLYQELH 320  
229 QY NMFMEDAETHGIEHNFCRNPDADKPCFKIVTNDKVWEYCDVSACS--QDVAYPEE 286  
321 Db VDSVGAALLGLGHAYCRNPDNERPWCYV-VKDSALSWCYCLEACESLTRVQLSPDL 379  
287 QY SPTPESTKLPDFDSCGKTEIAERKIK-RIYGGFKSTAGKHWPQASLOSSPLTISMPCQH 345  
380 Db LATLEPASPRQACGRHKKRTFLRPRIIGSSSLPGSHPWLAIVIG-----DS 430  
346 QY FCGGALIHPCWVLTAAHCTDIKTRH--LKVVLGQDLKKEFHQSRAVEKIFKYSYHNE 403  
431 Db FCAGSLVHTCWVYSAAHCFSHSPRDSVSVVLGQHFNRRTDVTQTFGIEKIYIPTYLSV 490  
404 QY RDEIPHNDIALKLPVGDHCHCALESKYKVTCLPD--GSFPSSGSECHISGWG-VTERGKG 460  
491 Db FNPDSH-DLVLRILKKGDRCATRSQFVQPCLEPEPGSTFFAGHKCQIAGWGLDENVSG 549  
461 QY -SRQLLDAKVLIANTLNSRLQSDHMDMSICAGNLQKPGQDTCQDSDGGPPTCEKDG 519  
550 Db YSSSLREALVPLVADHKCSPEVYAGDISPNMLCAGYFDCK-SDACQDSDGGPLACEKNG 608  
520 QY TTYVYGVISWGLECKK--RPGVYTVTKFLNWKATIK 555  
609 Db VAYLYGISWGDGCGRLHKPGVITRVANYVDNDRIR 646

RESULT 4  
S28941  
coagulation factor XIIa (EC 3.4.21.38) - guinea pig (fragment)  
N:Alternate names: Hageman factor  
C:Species: *Cavia porcellus* (guinea pig)  
C>Date: 25-Feb-1994 #sequence\_revision 03-Aug-1995 #text\_change 21-Jan-2000  
C:Accession: S28941  
R:Semba, U.; Yamamoto, T.; Kunisada, T.; Shibuya, Y.; Tanase, S.; Kambara, T.; Okabe, H.  
Biochim. Biophys. Acta 1159, 113-121, 1992  
A:Title: Primary structure of guinea-pig Hageman factor: sequence around the cleavage site  
A:Reference number: S28941; MUID:93003367; PMID:1390917  
A:Accession: S28941  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-603 <SEM>  
A:Cross-references: EMBL:X68615; NID:G49578; PIDN:CAA48600.1; PID:G49579  
C:Superfamily: coagulation factor XII; EGF homology; fibronectin type I repeat homology  
C:Keywords: hydrolase; coagulation factor XII; EGF homology; fibronectin type I repeat homology  
F:46-87/Domain: fibronectin type II repeat homology <F12>  
F:134-169/Domain: fibronectin type I repeat homology <F1>  
F:177-208/Domain: EGF homology <EGF>  
F:216-294/Domain: kringle homology <KR>  
F:359-597/Domain: trypsin homology <TRY>

Query Match 24.5%; Score 773.5; DB 2; Length 603;  
Best Local Similarity 34.1%; Pred. No. 1e-40;  
Matches 208; Conservative 74; Mismatches 205; Indels 123; Gaps 29;

QY 17 VGKTACGFLMSLESLLDPDWTQDQYSDYEDYNQBEEN-----TSSTLTHAENPDWY 69  
Db 37 VGLTVTG-----EPCYFFPQYNRLYHCHIKGRGPRPWCATTNFPDQDQWAY 86  
QY 70 ---TEDQADPC-QPNPCEHGDCILVHGSTFTCSCLAPFSGNKKCKQVQNTCKNDKPNCGRQCL 125  
Db 87 CLBPKVKVDHCSKHNPQCRGGICVNTLLSPHCLCPDHLTGKHCQ-----REKC 134

126 QY LITQSPYY-----RCVCKHPVTGSCSQV--PVCRRNPONGATCSRHK 169  
135 Db FEPQJHRRFFHENEIWRIGPAGVAKCHKGP--DAHCKQMSQSQQTNPCLNGGRCL--E 190  
170 QY RRSKFTACPDQFGKFCFCEIGSD- DCYVGDGYSYRGKRNRTVNOHACLYNWSHLLQENY 228  
191 Db VEGHILCDPMGTGTPFCDDLTDTASCYBGRGYSYRGMARITTVSGAKCORWASEATYR--- 247  
229 QY NMFMEDAETHGIEHNFCRNPDADKPCFKIVTNDKVWEYCDVSACS--AQDVAYPEE- 285  
248 Db NMTAEQALRRGLGHTFCRNPNDTRPWCFTVMGN-RLSWEYCDLAQCYPPQPTATPHD 306  
286 QY --ESPTPESTK-----LPGFDS--CGKTEIAER--KIKRIYGGPKS 320  
307 Db REHPKLPSSRLSILQTPQTTQNALANELPETSSLLCGQ-RLRKRLLSSLSRIYVGLVA 365  
321 QY TAGKHWPQASLOSSPLTISMPCQHFCGALHPCWVLTAAHCTDIK--TRHLKVLGQD 378  
366 Db LPGAEPYIAALYWG-----SNFCSGLIAPCWLTAAHCLQNRPAPELKVVLGQD 416  
379 QY DLKKEBFHQ-----SFRVEKIFKISHYNERDEIPHNDIALKL-KPVDGHCHCALESKYK 432  
417 Db RHNSCEHCQTAVHSHYRLHEAFSPSSY-----LNDLALLRLQKSDAGSCAQLSPYVQ 469  
433 QY TVCLPDG-SFPSSGSE---CHISGWGTETG--KGRQLLDAKVLIANTLNSRLQSDHMD 486  
470 Db TVCLPSGAPPSESETTCCEVAGWGHQFEGAEYSFLOEAQVPLISSRCSPEVHGDA 529  
487 QY IDDSMLCAGNLQKPGQDTCQDSDGGPPTCEKDGTYV---VYGVISWGLECKK--RPGVY 541  
530 Db FLGMLCAQFLE--GGTDACQDSDGGPLVCEDEAAEHLRLGIVSWGSGCGDRNKGVT 588  
542 QY QVTKFLNWK 551  
589 Db DVASYLTWIQ 598

RESULT 5  
UKHUT  
t-plasminogen activator (EC 3.4.21.68) precursor [validated] - human  
N:Alternate names: t-PA; tissue plasminogen activator  
C:Species: *Homo sapiens* (man)  
C>Date: 14-Nov-1983 #sequence\_revision 14-Nov-1983 #text\_change 08-Dec-2000  
C:Accession: A94004; A23529; J70562; A93293; S02125; A91343; A93951; A91322; A54645; I6C  
R:NY, T.; Eigh, F.; Lund, B.  
Proc. Natl. Acad. Sci. U.S.A. 81, 5355-5359, 1984  
A:Title: The structure of the human tissue-type plasminogen activator gene: correlation  
A:Reference number: A94004; MUID:84298137; PMID:6089198  
A:Accession: A94004  
A:Molecule type: DNA  
A:Residues: 1-562 <NVT>  
A:Cross-references: GB:I00141  
A>Note: the codon given for residue 93 (ACC) is inconsistent with the authors' translation  
R:Frieznher Degen, S.J.; Rajput, B.; Reich, E.  
J. Biol. Chem. 261, 6972-6985, 1986  
A:Title: The human tissue plasminogen activator gene.  
A:Reference number: A23529; MUID:86196143; PMID:3009482  
A:Accession: A23529  
A:Molecule type: DNA  
A:Residues: 1-562 <DEG>  
A:Cross-references: GB:X03021; NID:G339817; PIDN:AAA98809.1; PID:G339818  
R:Itagaki, Y.; Yasuda, H.; Morinaga, T.; Mitsuda, S.; Higashio, K.  
Agric. Biol. Chem. 55, 1225-1232, 1991  
A:Title: Purification and characterization of tissue plasminogen activator secreted by  
A:Reference number: J70562; MUID:91291340; PMID:1368681  
A:Accession: J70562  
A:Molecule type: mRNA  
A:Residues: 31-562 <ITA>  
A:Cross-references: DDBJ:D01096; NID:G220128; PIDN:BAA00881.1; PID:G441174  
A:Experimental source: embryonic lung fibroblast IWR-90 cells  
A>Note: part of this sequence, including the amino end of the mature protein, was confirmed  
R:Pernica, D.; Holmes, W.E.; Kohr, W.J.; Harkins, R.N.; Vehar, G.A.; Ward, C.A.; Bennett

Nature 301, 214-221, 1983  
A:Title: Cloning and expression of human tissue-type plasminogen activator cDNA in Esche  
A:Reference number: A93293; MUID:83115262; PMID:6337343  
A:Accession: A93293  
A:Molecule type: mRNA  
A:Residues: 1-562 <PEN>  
A:Cross-references: GB:100141  
A:Experimental source: melanoma cells  
R:Sasaki, H.; Saito, Y.; Hayashi, M.; Otsuka, K.; Niwa, M.  
Nucleic Acids Res. 16, 5695, 1988  
A:Title: Nucleotide sequence of the tissue-type plasminogen activator cDNA from human fe  
A:Reference number: S02125; MUID:88262579; PMID:3133640  
A:Accession: S02125  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-562 <SAS>  
A:Cross-references: EMBL:X07393; NID:G37243; PIDN:CAA30302.1; PID:G37244  
A:Experimental source: fetal lung cells  
R:Kagiguchi, H.; Tagawa, M.; Hatanaka, K.; Ikari, T.; Saito, A.; Bando, H.; Okada, K.; Ma  
FEBS Lett. 189, 145-149, 1985  
A:Title: Expression in Escherichia coli of finger-domain lacking tissue-type plasminogen  
A:Reference number: A91343; MUID:85285620; PMID:3896853  
A:Accession: A91343  
A:Molecule type: mRNA  
A:Residues: 1-38, 'G', 86-433, 'E', 435-562 <KAG>  
A:Experimental source: Detroit 562 cells; ATCC 138  
R:Exlund, T.; Ny, T.; Ranby, M.; Heden, L.O.; Palm, G.; Holmgren, E.; Josephson, S.  
Proc. Natl. Acad. Sci. U.S.A. 80, 349-352, 1983  
A:Title: Isolation of cDNA sequences coding for a part of human tissue plasminogen activ  
A:Reference number: A93951; MUID:83169656; PMID:6572897  
A:Accession: A93951  
A:Molecule type: mRNA  
A:Residues: 251-358 <EDI>  
A:Experimental source: melanoma cells  
R:Poehl, G.; Kallstrom, M.; Bergsdorf, N.; Wallen, P.; Jornvall, H.  
Biochemistry 23, 3701-3707, 1984  
A:Title: Tissue plasminogen activator: peptide analyses confirm an indirectly derived an  
differences.  
A:Reference number: A90488; MUID:85000468; PMID:6433976  
R:Poehl, G.; Kaplan, L.; Einarsson, M.; Wallen, P.; Jornvall, H.  
FEBS Lett. 168, 29-32, 1984  
A:Title: Differences between uterine and melanoma forms of tissue plasminogen activator.  
A:Reference number: A91322; MUID:84158956; PMID:6538514  
A:Accession: A91322  
A:Molecule type: protein  
A:Residues: 33-45/311-320 <POH>  
A:Experimental source: uterus  
R:van Zonneveld, A.J.; Veerman, H.; Pannekoek, H.  
J. Biol. Chem. 261, 14214-14218, 1986  
A:Reference number: A37567; MUID:87033611; PMID:3021732  
R:Verheijen, J.H.; Caspers, M.P.M.; Chang, G.T.G.; de Munk, G.A.W.; Pouwels, P.H.; Enger  
EMBO J. 5, 3525-3530, 1986  
A:Title: Involvement of finger domain and kringle 2 domain of tissue-type plasminogen ac  
A:Reference number: A37568; MUID:87161761; PMID:3030730  
R:Dodd, I.; Nunn, B.; Robinson, J.H.  
Thromb. Haemost. 59, 523-528, 1988  
A:Title: Isolation, identification and pharmacokinetic properties of human tissue-type P  
A:Reference number: A60922; MUID:89044681; PMID:3142086  
R:Harris, T.J.R.; Patel, T.; Marston, F.A.O.; Little, S.; Emtage, J.S.; Odenakker, G.;  
Vol. Biol. Med. 3, 279-292, 1986  
A:Title: Cloning of cDNA coding for human tissue-type plasminogen activator and its exp  
A:Reference number: A54645; MUID:86284200; PMID:3030401  
A:Accession: A54645  
A:Molecule type: mRNA  
A:Residues: 1-562 <HAP>  
A:Cross-references: GB:M5518; NID:G190031; PIDN:AA6011.1; PID:G190032  
A:Note: parts of this sequence were confirmed by peptide sequencing  
R:Reddy, V.B.; Garraffo, A.J.; Sasak, H.; Wei, C.

DNA 6, 461-472, 1987  
A:Title: Expression of human uterine tissue-type plasminogen activator in mouse cells u  
A:Reference number: 160110; MUID:88054470; PMID:2824147  
A:Accession: 160110  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-562 <RES>  
A:Cross-references: GB:M18182; NID:G340176; PIDN:AAA36800.1; PID:G340177  
R:Fisher, R.; Waller, E.K.; Grossi, G.; Thompson, D.; Tizard, R.; Schleuning, W.D.  
J. Biol. Chem. 260, 11223-11230, 1985  
A:Title: Isolation and characterization of the human tissue-type plasminogen activator  
A:Reference number: 155232; MUID:85289338; PMID:3161893  
A:Accession: 155232  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-36 <RE2>  
A:Cross-references: GB:M11890; NID:G339837; PIDN:AAA61213.1; PID:G339839  
C:Comment: Cleavage by plasmin or trypsin produces two chains held together by a single  
C:Comment: t-PA converts plasminogen to plasmin by hydrolyzing a single Arg-Val bond. I  
C:Comment: t-PA binds chain A of fibrin by kringle 2 and the fibronectin type I repeat.  
C:Genetics:  
A:Gene: GDB:PLAT  
A:Cross-references: GDB:119496; OMIM:173370  
A:Map position: 8p12-8p12  
A:Introns: 24/3; 39/1; 85/1; 122/1; 180/2; 211/1; 268/2; 297/1; 362/2; 408/1; 455/3; 51  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat ho  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; plasma; serine proteinase  
F1-23/Domain: signal sequence #status predicted <SIG>  
F24-32/Domain: propeptide #status predicted <PRO>  
F33-562/Product: t-plasminogen activator #status experimental <NAT>  
F33-310/Product: t-plasminogen activator chain A #status experimental <ACH>  
F41-78/Domain: fibronectin type I repeat homology <1F1>  
F46-119/Domain: EGF homology <EGF>  
F127-208/Domain: kringle homology <K1>  
F215-296/Domain: kringle homology <KR2>  
F311-562/Product: t-plasminogen activator chain B #status experimental <BCH>  
F41-71,69-78,86-97,91-108,110-119,127-208,148-190,179-203,215-296,236-278,267-291,299-  
F152,483/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F219/Binding site: carbohydrate (Asn) (covalent) (partial) #status experimental  
F310-311/Cleavage site: Arg-Ile (plasmin, trypsin) #status experimental  
F357,406/Active site: His, Asp #status predicted  
F513/Active site: Ser #status experimental  
Query Match 24.3%; Score 766.5; DB 1; Length 562;  
Best Local Similarity 35.0%; Pred. No. 2.6e-40;  
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;  
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKCCQ-KVONTCKDNP--CGRG----- 123  
DB 86 CSEPCFNGTCQQAALYFDFVCCPEGFAGKCCCEITRATCYEDQISYRGTTSTAESG 145  
QY 124 -QCLITQSPYRYRCVCKHPYTGSCSQVVPVCRENPQ-----NGATCSRHKRRSKFTQAC 178  
DB 146 AECTNWN-----SALAQKPYSGR-----RDAIRLGLGHNHCRNPRDSDKPCWCV 192  
QY 179 PDQFK-GKF-----CEIGSDDCYVDGYSYRGKNTVQHQACLYNNSHLLLOENY 228  
DB 193 ---FKAGKYSSEFCSTPACSEGNDSYFNGSGATRGTHSLTESGASCLPWNMILIGKV 249  
QY 229 NMFMEDAETHGIGEHNFNCRNPDADKPMCFIKVNDKVMKVCYDSACSAQDVAYPEESP 288  
DB 250 TQNPESAQALGLGKHNCRNPDGAKPWCHV-LKNRRLTWEYCDVPCS----- 297  
QY 289 TEPSTKLPFGSCGKGTETAEKRIYCGFKSTAGKHPWQASLOSLPLTTISMQGHFCG 348  
DB 298 -----TCGLRQYSQPF-RIKGLFADTASHPWQAIFAK--HRSKDPGRFLCG 343  
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVVGLGQDLKKEEFHQSRFEVEKIFKYSHNERDE 406  
DB 344 GILISCSWLSAAHCFQSRFPFPHLTVILGTYRVVPCGEEQKFEVEKYVHKSPD--DD 401  
QY 407 IPNDIALILKLPVDGHCALSKYKVTVCPLP--DGSFFSGSECHISGWGTETGKG--SR 462

Db 402 TYNDIALQLKSDSSRCAQESSVVRVCLPPADLQLPDWTECELSGYGKHEALSPYSE 461  
Qy 463 QLLDAKVLIANTICNRQLYDHMDSDMICAGNLQKPG-----QDTCQSDSGSLPTCEK 517  
Db 462 RLKEAHLVRLYSPSRCTSOHLNRTVTDNMLCAGTRSGGQFQANJHDACQSDSGSLVCLN 521  
Qy 518 DGTVVYGVISWGLECGKR--PGVTVQTKFLNWKATIK 555  
Db 522 DGRMTLVGIIISGLCCGQKQVPGYTKVNTYLDNR 561  
RESULT 6  
A35029  
t-plasminogen activator (EC 3.4.21.68) precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A35029; A31597  
R:feng, P.; Ohlsson, M.; Ny, T.  
J. Biol. Chem. 265, 2022-2027, 1990  
A:Title: The structure of the TATA-less rat tissue-type plasminogen activator gene. Spec  
A:Reference number: A35029; MUID:90130448; PMID:2105315  
A:Accession: A35029  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-559 <FEN>  
A:Cross-references: GB:M31197; NID:G207429; PIDN:AAA42261.1; PID:G207431; GB:J05226  
R:NY, T.; Leonardson, G.; Hsueh, A.J.W.  
DNA 7, 671-677, 1988  
A:Title: Cloning and characterization of a cDNA for rat tissue-type plasminogen activator  
A:Reference number: A31597; MUID:89170114; PMID:3148445  
A:Accession: A31597  
A:Molecule type: mRNA  
A:Residues: 1-379, 'K', 381-559 <NYT>  
A:Cross-references: GB:M23697; NID:G530159; PIDN:AAA41812.1; PID:G530160  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:18-29/Domain: propeptide #status predicted <PRO>  
F:30-559/Product: t-plasminogen activator #status predicted <MAT>  
F:308-309/Cleavage site: Arg-Ile (plasmin, trypsin) #status predicted <ACH>  
F:355,404,510/Active site: His, Asp, Ser #status predicted <I1>  
F:83-116/Domain: fibronectin type I repeat homology <I1>  
F:124-205/Domain: kringle homology <KR1>  
F:213-294/Domain: kringle homology <KR2>  
F:303-559/Product: t-plasminogen activator chain B #status predicted <BCH>  
F:308-68,66-75,83-94,88-105,107-116,124-205,145-187,176-200,213-294,234-276,265-289,297-4  
F:145,481/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:308-309/Cleavage site: Arg-Ile (plasmin, trypsin) #status predicted  
F:355,404,510/Active site: His, Asp, Ser #status predicted

Query Match 24.0%; Score 757.5; DB 1; Length 559;  
Best Local Similarity 32.5%; Pred. No. 9,4e-40;  
Matches 202; Conservative 86; Mismatches 192; Indels 141; Gaps 29;  
Qy 8 LHVLLMALVGTACG--FSLMSLLESPLDPDTPQDYD-----SYEDYNGEENTSS 57  
Db 6 LCVLL-----CGVAFTL-----PDQIHRRFRGARSYRATCRDEQTQT 45  
Qy 58 ILTHAENPDW---YTEQADQCQNP-----CEHGDC--LVHGSTF 95  
Db 46 --TYQCHOSWLRPMLRGNRVEYCRNCNGLAOCHSVVRSCSEPRCFNGGTCCQALYPSDF 103  
Qy 96 TCSCLAIPSGNKKQ--KVQNTCKDNPCGRQCQLITQSPYRVCVKHPYTGSPC----- 147  
Db 104 VQCPDGFVGKCDIDTRATCFE---GGG---IT-----YRGTWSTAENGAEINWNSSA 152  
Qy 148 --SQVFCVRPNPCNGATCSHKRSKFTACPDQ-----FK-GKF-----C 187  
Db 153 LSQKPYSAARRPNAIKLG--LGNHN-----YCRNPRDVKPCWYFKAGKYTFECSTPAC 205  
Qy 188 EIG-SDDCVGVGDSYRGGMNRTVQACLYNWSHLLQLQENYMPMEDAETHGICEHNF 246

Db 206 PKGPTEDCVYGVYRGTHSFTTSKASCLPWNMILIKTWTAWRANSQALGLGRHNYC 265  
Qy 247 RUPDADERKWCPIKNTNDKVKWEYCDVSACSNQDVAYPEESTPESTKPLGPDFSDSGKTEI 306  
Db 266 RRPDGDGAPWCHV--NKDRKLTWEYCDMSPCS-----TCGLQKY 302  
Qy 307 AERKIKRIYGVGFKSTAGKHPMQASLQSSILPLTISMPQGHFCGALIHFCWVLTAAHC--T 364  
Db 303 KQPQF-RIKGGFTDITSHPMQAAL--FVGNKRSFGRFLCGGVLISSCWVLSAAHCFVE 359  
Qy 365 DIKTRHLKVLGGDLKKEEFHEQSRVEKFKYSHYNERDEIPNDTALLKLPVDGHC 424  
Db 360 RFPFPHLKVVLGRTYRVVPGEEQTFEIKYIVHKEFD--DDTYDNDLALQLGRSDSSOC 417  
Qy 425 ALESKYKVTVCIPDG--SPPSGSECHISQWGVYETGKG---SQLLDAKVLIANTICNRS 480  
Db 418 AQESSVGTACLPDQVQLPDWTECELSGYGKHEASSPFFSDRLKEAHLVRLYSPSRCTSQ 477  
Qy 481 QYDHMDSDMICAGNLQKPG-----QDTCQSDSGSLPTCEKDTYVYGVISWGLECGKR 536  
Db 478 HLFNKTITSNMLCAGDTRTGNQDHYDACQSDSGSLVCMIDKRWMTLLGIISWGLGCGOK 537  
Qy 537 --PGVTVQTKFLNWKATIK 555  
Db 538 DYPGIYTKVNTYLDNR 558  
RESULT 7  
A29941  
t-plasminogen activator (EC 3.4.21.68) precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A29941; S48205; S48207; S48206  
J:Rickles, R.J.; Darrow, A.L.; Strickland, S.  
J. Biol. Chem. 263, 1563-1569, 1988  
A:Title: Molecular cloning of complementary DNA to mouse tissue plasminogen activator mR  
A:Reference number: A29941; MUID:88087303; PMID:2826484  
A:Accession: A29941  
A:Molecule type: mRNA  
A:Residues: 1-559 <R1C>  
A:Cross-references: GB:J03520; NID:G202109; PIDN:AAA40470.1; PID:G202110  
R:Lijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.  
Eur. J. Biochem. 224, 863-871, 1994  
A:Title: Characterization of the murine plasma fibrinolytic system.  
A:Reference number: S48202; MUID:95010076; PMID:7523120  
A:Accession: S48205  
A:Molecule type: protein  
A:Residues: 33-37, 'X', 39-40 <LIJ>  
A:Accession: S48207  
A:Molecule type: protein  
A:Residues: 309-316 <LI2>  
A:Accession: S48206  
A:Molecule type: protein  
A:Residues: 33-37, 'X', 39-40 <LIW>  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:18-29/Domain: propeptide #status predicted <PRO>  
F:30-559/Product: t-plasminogen activator #status predicted <MAT>  
F:308-309/Product: t-plasminogen activator chain A #status predicted <ACH>  
F:38-75/Domain: fibronectin type I repeat homology <I1>  
F:83-116/Domain: EGF homology <EGF>  
F:124-205/Domain: kringle homology <KR1>  
F:213-294/Domain: kringle homology <KR2>  
F:303-559/Product: t-plasminogen activator chain B #status predicted <BCH>  
F:308-68,66-75,83-94,88-105,107-116,124-205,145-187,176-200,213-294,234-276,265-289,297-4  
F:149,481/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:308-309/Cleavage site: Arg-Ile (plasmin, trypsin) #status predicted  
F:355,404,510/Active site: His, Asp, Ser #status predicted

Query Match 23.6%; Score 745.5; DB 1; Length 559;

Best Local Similarity 35.0%; Pred. No. 5.2e-39;  
Matches 183; Conservative 76; Mismatches 173; Indels 91; Gaps 21;

QY 77 COPNCEHGGDC--LVHGSTFTCSCLAFPSGNKCO-KVONTCKONPCGRGCLITQSPPY 133  
Db :  
QY 83 CSEPRCFNGGTCCQALYSDFVCQCPDGFVKRCRCDIDTRATCFEE---QG---IT----- 131  
Db :  
QY 134 YRCVCKHPTGPSC---SQVPV-----CRPNPQC-----NGATCSRHKRSKFTC----- 176  
Db :  
QY 132 YRGTWSTAESGAACINWNSSVLSPXYVARRPNAIKLGLHNHYCRNPDRDLKPWCYVFK 191  
Db :  
QY 177 -----ACPDQPKGFCPEGSDCYVDGDYSYRGKKNRVTNQHAQLYNSHLLL 224  
Db :  
QY 192 AGKYTFTEFCSTPACE---KGK-----SEDYVGKGVTVRGTHSLTTSOASCLPWNSIVLM 243  
Db :  
QY 225 DENTNMFMEDAEITHIGBHFNCRNPDADBKPCFFIKVFNDKVKVEYCDVSACSADVAYP 284  
Db :  
QY 244 GKSYTAWTNTSQAALGARHNYCRNFDGARPHCHV-MKDRKLTYEYCDMSPCS----- 295  
Db :  
QY 285 BESPEPTSKLPFGPSCGKTEIAERKIIRIYGFKSTAGKIPWQASLOSSLPLTIMPOG 344  
Db :  
QY 296 -----TCGLRQY-KRFQPRIKGLYTDTISHPWOAAI--FVNKRKSPOER 337  
Db :  
QY 345 HFCGGALIHPCWVLTAACH--TDIKTRLKVVLGQDLKKERFHEQSFRVEKIFKYSHYN 402  
Db :  
QY 338 FLCGGVLSSCWVLSAACHFLBRFPNLKVVLGRTRYVVFGEETFEIEKYIVHBEPD 397  
Db :  
QY 403 ERDEIPHNDIALLKXPDVGHCALESKYVTKVLCLDGS--FPSGECHISGWGTETGKG 460  
Db :  
QY 398 ---DDTYDNDAIALQRSOSKQCAQSSSVGTACLDPNLQLPDMTECELSGYSGKEARSP 455  
Db :  
QY 461 ---SRQLLAKVKLIANTLNCNSRLQYDHMDIDSAMI CAGNLQKPG----QDTCGDSGGPLT 514  
Db :  
QY 456 PFSDRLKEAHVELYPSRCTSCHLENNKTVTNNMLCAGDTRSGGNODLHDACQDSGGPLV 515  
Db :  
QY 515 CEKDGTYYVGIIVSWGLEGCKR--PGVIYQTQVTKFLNWIKATIK 555  
Db :  
QY 516 CMNKQMTLTGIIISGLGCGGXDPGVYTKVNYLDWIHDNMK 558  
Db :

RESULT 8  
KFHU12  
coagulation factor XIIa (EC 3.4.21.38) precursor [validated] - human  
NAlternate names: Hageman factor (activated)  
CSpecies: Homo sapiens (man)  
CDate: 27-Nov-1985 #sequence revision 30-Jun-1991 #text\_change 08-Dec-2000  
CAccession: A29411; A26814; A00930; A25191; A22248; A21037  
R;Triposdi, M.; Citarella, F.; Guida, S.; Galeffi, P.; Fantoni, A.; Cortese, R.  
Nucleic Acids Res. 14, 3146, 1986  
R;Cool, D.E.; Edgell, C.J.S.; Louie, G.V.; Zoller, M.J.; Brayer, G.D.; MacGillivray, R.T.A.  
J. Biol. Chem. 260, 13666-13676, 1985  
J. Biol. Chem. 262, 13662-13673, 1987  
A>Title: Characterization of the human blood coagulation factor XII gene. Intror/exon ge  
A;Reference number: A29411; MUID:88007593; PMID:2688762  
A;Accession: A29411  
A:Molecule type: DNA  
A;Residues: 1-615 <CO>  
A;Cross-references: GB:M17466; GB:J02807; NID:g180355; PIDN:AAB59490.1; PID:g180357  
R;Triposdi, M.; Citarella, F.; Guida, S.; Galeffi, P.; Fantoni, A.; Cortese, R.  
Nucleic Acids Res. 14, 3146, 1986  
A>Title: cDNA sequence coding for human coagulation factor XII (Hageman).  
A;Reference number: A26814; MUID:86176794; PMID:3754331  
A;Accession: A26814  
A:Molecule type: mRNA  
A;Residues: 4-615 <TRI>  
A;Cross-references: GB:M31315; NID:g182291; PIDN:AAA70225.1; PID:g182292  
R;Cool, D.E.; Edgell, C.J.S.; Louie, G.V.; Zoller, M.J.; Brayer, G.D.; MacGillivray, R.T.  
J. Biol. Chem. 260, 13666-13676, 1985  
A>Title: Characterization of human blood coagulation factor XII cDNA. Prediction of the  
A;Reference number: A00930; MUID:86033830; PMID:3877053  
A;Accession: A00930  
A:Molecule type: mRNA  
A;Residues: 14-332, S\_334-615 <CO2>  
A;Cross-references: GB:M11723; NID:g180358; PIDN:AAA51986.1; PID:g180359  
R;Que, B.G.; Davie, E.W.  
Biochemistry 25, 1525-1528, 1986

A>Title: Characterization of a cDNA coding for human factor XII (Hageman factor) .  
A:Reference number: A25191; MUID:86216049; PMID:3011063  
A:Accession: A25191  
A:Molecule type: mRNA  
A:Residues: 146-378, 'G', 380-615 <QUB>  
A:Cross-references: GB:M13147; NID:G180360; PIDN:AAA0224.1; PID:G180361  
R:McMullen, B.A.; Fujikawa, K.  
J. Biol. Chem. 260, 5328-5341, 1985  
A>Title: Amino acid sequence of the heavy chain of human alpha-factor XIIIa (activated H  
A:Reference number: A22248; MUID:185182674; PMID:3986654  
A:Accession: A22248  
A:Molecule type: protein  
A:Residues: 20-379 <MCW>  
R:Fujikawa, K.; McMullen, B.A.  
J. Biol. Chem. 258, 10924-10933, 1983  
A>Title: Amino acid sequence of human beta-factor XIIa.  
A:Reference number: A21037; MUID:83291041; PMID:6604055  
A:Accession: A21037  
A:Molecule type: protein  
A:Residues: 354-362;373-615 <FUG>  
R:R.Harris, R.J.; Ling, V.T.; Spellman, M.W.  
J. Biol. Chem. 267, 5102-5107, 1992  
A>Title: O-linked fucose is present in the first epidermal growth factor domain of factor  
A:Reference number: A4606; MUID:92184750; PMID:1544894  
A:Contents: annotation, carbohydrate binding site  
C:Genetics:  
A:Gene: GDB:F12  
A:Cross-references: GDB:119892; OMIM:234000  
A:Map position: 5q34-5qter  
A:Intons: 19/3; 39/1; 72/2; 96/1; 133/1; 177/1; 212/1; 267/2; 340/1; 417/2; 463/1; 511/  
C:Complex: factor XII, prekallikrein, and HMW kininogen form a complex bound to anionic  
C:Function:  
A:Description: factor XIIIa catalyzes the proteolytic activation of plasminogen, plasma i  
ikrein  
A:Pathway: blood coagulation; fibrinolysis  
C:Superfamily: coagulation factor XII; EGF homology; fibronectin type I repeat homology  
C:Keywords: blood coagulation; fibrinolysis; glycoprotein; hydrolase; kringle; plasma;  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-372/Domain: signal sequence #status predicted <SIG>  
F:47-88/Domain: fibronectin type II repeat homology <FB2>  
F:98-130/Domain: EGF homology <EG1>  
F:135-170/Domain: fibronectin type I repeat homology <1F1>  
F:178-209/Domain: EGF homology <EG2>  
F:217-295/Domain: kringle homology <KR>  
F:298-356/Region: proline-rich  
F:354-362/Region: coagulation factor XIIIa, beta form #status experimental <B12  
F:373-609/Domain: trypsin homology <TRY>  
F:98-110,104,119,121-130,135-163,161-170,178-189,183-198,200-209,217-295,238-277,266-299  
F:109/Binding site: carbohydrate (Thr) (covalent) #status experimental  
F:249,433/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:299,305,328,329,337/Binding site: carbohydrate (Thr) (covalent) #status predicted  
F:308/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F:412,461,563/Active site: His, Asp, Ser #status predicted

```

Query Match      23.5%; Score 741; DB 1; Length 615;
Best Local Similarity 31.3%; Pred. No. 1.le-38;
Matches 206; Conservative 79; Mismatches 210; Indels 164; Gaps 34;

Qy      8 LHVLLMALVGKTACGFLSLSLESLD-PDM-TPDQYDYSYED-----YN 50
      : |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db      1 MRALLLL-----GFLVLSLESLISPPWEAPKEHKYKAEHTVWLVTGCEPHFPQ 52

Qy     51 QEENTSGSTTHAENP---DWYVT-----EDQ-----ADQC-QENPCEHGGDCLV 90
      : |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db     53 YHRQLYHKTHKGRPGQFWCAATTNFDDQDQWGYCLEPKVKYDCKSHSPCKGKGTGVN 112

Qy     91 HGSFTFTCSCLAPFSGNCKQKVQNTCKDNPCGRGQCL-----ITOSPYPYRCVKKH 140
      : |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db    113 MPSPGPHCLCPQLHTGNCHQK-----EKCPEPQLLRPPHKNEIYWYRTEQAQAVARCQCKG 165

Qy    141 PYTGPSCSQV--PVCPNPNCQNGATC---SRHKRSKFTCAQDPDFKGFCEIGSD--DCY 195
      : ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db    166 P--DAPHCQLAQAACRTNPLCHGRCLEVEGHR-----LCHQPVGTGTFCDVDTKASY 218

```





```
QY 400 HYNREDEIPNDIALLKLPVDGCHCALESKYKVTVCPLPDGS--FPSGSECHISGVTET 457
Db 312 EFD--DDTYNDNALLQLKSGSQCAQSDSVRAICLPEANLQLPDWTSCELSGYGGKHS 369
QY 458 GKG--SRQLDAKYKIANTLNSRQLYDHMDSDMCAQNLQK-----PG-QDTQGDSDG 510
Db 370 SSPFYBQLKEGHVRLYPPSRCTSKFLNKTNTNNMLCAGDTRSGBIYFNVDACQGDSDG 429
QY 511 GPITCEKDGYYVYVGVSMGLECGKR--PGVYTVTKFNLNWKATIK 555
Db 430 GPLVCMNDNMTLLGIISWVGCGEKDIPGVYTKVNYLWGRDNR 476

RESULT 11
JS0599
t-plasminogen activator (EC 3.4.21.68) beta precursor - common vampire bat
N:Alternate names: tissue plasminogen activator
C:Species: Desmodus rotundus (common vampire bat)
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: JS0599
R:Kraetzschmar, J.; Haendler, B.; Langer, G.; Boidol, W.; Bringmann, P.; Alagon, A.; Don
Gene 105, 229-237, 1991
A:Title: The plasminogen activator family from the salivary gland of the vampire bat Des
A:Reference number: JS0597; MUID:92039036; PMID:1937019
A:Accession: JS0599
A:Molecule type: mRNA
A:Residues: 1-431 <KRA>
A:Cross-references: GB:M63989; NID:G166076; PID:AAA1594.1; PID:G166077
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-36/Domain: propeptide #status predicted <PRO>
F:37-43/Product: plasminogen activator beta #status predicted <PLA>
F:41-74/Domain: EGF homology <EGF>
F:82-163/Domain: kringle homology <KR>
F:180-425/Domain: trypsin homology <TRY>
F:41-52,46-63,65-74,82-163,103-145,134-158,168-299,211-227,219-288,313-388/Disulfide bon
F:139,352/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:179-180/cleavage site: His-Ser (plasmin)
F:226,275,382/Active site: His, Asp, Ser #status predicted
F:345-361,378-406/Disulfide bonds: #status predicted

Query Match 22.8%; Score 720.5; DB 2; Length 431;
Best Local Similarity 36.0%; Pred. No. 1.4e-37;
Matches 157; Conservative 73; Mismatches 159; Indels 47; Gaps 13;

QY 134 YRCVKHPYTGPSQSVVPCRPNCONGATCSRHKRSKFTACPPQFGKTCETGSD- 192
Db 28 YRQLARGSRAYGGCSEL-----RCFNGGTCWAASFDVCCQPKGYTGKQCEVDTHA 80
QY 193 DCYVGDGYSYRGKMRNTVNOHACLYWNSHLLLOENYMFMEDETHGIGHNFCRNPAD 252
Db 81 TCYKDGQVYTRGTWSTESGAQCINWNSNLLTRTYNGRRSDAITLGLGNHNYCRNPDNN 140
QY 253 EKPWCFTKVTNDKWKVEYCDVSASQAQVAYPESPTSTKLPFGFSCGKTEIAERKIK 312
Db 141 SKPWCYV-IKASKETILEFCSPVPCSKA-----TCGLRKYKEPOLH 179
QY 313 RIYGFKSTAGHPQWASLQSLPLTISMPQGHFCGGALIHPCVWLTAACHCTDIK--TRH 370
Db 180 ST-GGLFTDITSHPQWAAIPAQ--NRRSSGERFLCGGLISSCWVLTAAHCFQERYPPQH 236
QY 371 LKVLGDQDLKKEBHEHQSFRVEKIFYKSHYNERDEIPNDIALLKLPVDGCHCALESKY 430
Db 237 LRVVLGRTYRVKPGKEEQTEFEKCIHIEEPD--DDTYNDNALLQLKSGSQCAQSDS 294
QY 431 VKTVCLPDGS--FPSGSECHISGVTETGKG--SRQLDAKYKIANTLNSRQLYDHMD 486
Db 295 VRAICLPEANLQLPDWTSCELSGYGGKHSPPFSEQKEGHVRLYPPSRCTSKFLNKT 354
QY 487 IDDSMICAGNLQK-----PG-QDTQGDSDGGPLTCEKDGYYVYVGVSMGLECGKR--PGV 539
Db 487 IDDSMICAGNLQK-----PG-QDTQGDSDGGPLTCEKDGYYVYVGVSMGLECGKR--PGV 539
```

```
Db 355 VTNNMLCAGDTRSGEIPYNVHDAQQDGGGLVCMNDNMTLLGIISWVGCGEKDIPGV 414
QY 540 YTVTKFNLNWKATIK 555
Db 415 YTKVNYLWGRDNR 430

RESULT 12
S45281
coagulation factor XIIa (EC 3.4.21.38) precursor - bovine (fragment)
N:Alternate names: Hageman factor (activated)
C:Species: Bos primigenius taurus (cattle)
C>Date: 10-Apr-1995 #sequence_revision 22-Apr-1995 #text_change 21-Jan-2000
C:Accession: S45281; A61329
R:Shibuya, Y.; Semba, U.; Okabe, H.; Kambara, T.; Yamamoto, T.
Biochim. Biophys. Acta 1206, 63-70, 1994
A:Title: Primary structure of bovine Hageman factor (blood coagulation factor XII): com
A:Reference number: S45281; MUID:94242782; PMID:8186251
A:Accession: S45281
A:Molecule type: mRNA
A:Residues: 1-593 <SHI>
A:Cross-references: GB:S70164
S:Note: the authors translated the codon GAG for residue 23 as Val, GAG for residue 70;
s Pro, CTC for residue 203 as Phe, GTG for residue 247 as Leu, CCG for residue 286 as G
is, and ATC for residue 505 as Leu
R:Fujikawa, K.; Walsh, K.A.; Davie, E.W.
Biochemistry 16, 2270-2278, 1977
A:Title: Isolation and characterization of bovine factor XII (Hageman factor).
A:Reference number: A61329; MUID:77181112; PMID:861210
A:Accession: A61329
A:Molecule type: protein
A:Residues: 10-16,'X',18-19,525-550 <FUJ>
C:Superfamily: coagulation factor XII; EGF homology; fibronectin type I repeat homology
C:Keywords: blood coagulation; fibrinolysis; glycoprotein; hydrolase; monomer; plasma;
F:37-78/Domain: fibronectin type II repeat homology <IF2>
F:98-120/Domain: EGF homology <EGF>
F:125-160/Domain: fibronectin type I repeat homology <FB1>
F:207-287/Domain: kringle homology <KR>
F:350-587/Domain: trypsin homology <TRY>
F:541/Active site: Ser #status predicted

Query Match 22.4%; Score 706.5; DB 2; Length 593;
Best Local Similarity 33.0%; Pred. No. 1.4e-36;
Matches 191; Conservative 69; Mismatches 183; Indels 135; Gaps 31;

QY 56 SSTLTHAENPDWYY---TEDQADPC-QNPCEHGGDCLVHGSTFTCSCLAFSGNKKQKV 111
Db 64 ATTFNFKDQWAWCLEPKVKVDCSKENPCQKGGTCVNMNPDGPRCICADHFTGKHKOK- 122
QY 112 QNTKDNPCGRQCCLITQSPVYR-----CVCCKHP--YTGPSCSQWV 151
Db 123 -----EKCF---EPOFFRFFHENEIWHLEPAGVYKCKGNAQCKPLASQ-- 166
QY 152 PVCPRNPONGATCSR---HKRSKFTCACPDQPKGKFCETG-SDDCY--VGDGYSYRGK 205
Db 167 -VCRINPCLNGSDCLQAEGHR-----LCRCAPSAGRLCDVDLKCASCYDDRDRLGSLYRGM 220
QY 206 MNRVTNQHACLYWNSHLLLOENYMFMEDETHGIGHNFCRNPADSKPCFIKVTNDK 265
Db 221 AGTTLGAPCQSWASEATY---WNVTAEVLNWLGDHAFCRNPNDTRPWCPI-WKGR 276
QY 266 VKWEYCDVSAC--SAQDVAYPEEP-----TEPSTKLPGFDS--CKTEIAE----- 308
Db 277 LSNYICRLAPCAAAGHEHFFPLPSALQKPBSTTQTPSLTSGWCSPTPLASGGPGGC 336
QY 309 -----RK-----IKRIYGGFKSTAGKHPQWASLOSPLTISMPQGHFCGGALHPCVWLTA 360
Db 337 GQRLRLKWLSSLRVVGGVLVAFGAHPYTAALYW-----DQHFCAAGSLIAPCVWLTA 387
QY 361 AHCTDIK--TRHLKVVVLGDQLKKEBHEHQS-----FRVEKIFK---YSHYNER 404
Db 388 AHCLQNPAPKELTVVLG-----QDRNQSCQEQCTQAVRDYRLHEAFSPITYQH---- 437
```

QY 405 DEIPNDIALKL-KPVDGCHCALESKYVTKVLCIPD-GSPFSGSE---CHISGWG-VTETG 458  
DB 438 -----DLALVRLQESADGCCAHPSPFPVQVCLPSTFAARPAESAAVCEVAGNWHQFEGG 491  
QY 459 KGSROLLDAKVKLIANTLCNRSQLYDHMDISMTICAGNLQKPGQDTCQGSQGGPLTCE-- 516  
DB 492 EYSSFLQEAQVPLIDPQRCSPADVHGAAFTQGMCLCAGFLE--GGTDAQCGDSGGPLVCEDE 550  
QY 517 -KDGYYVYGVISWGLECGKR--PGVYTVQVTKFLNWK 551  
DB 551 TPERQLILRGIVSWGCGNRLKPGVYTDVANYLAWIR 588  
RESULT 13  
J80597  
t-plasminogen activator (EC 3.4.21.68) alpha-1 precursor - common vampire bat  
N:Alternate names: tissue plasminogen activator  
C:Species: Desmodus rotundus (common vampire bat)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 16-Jul-1999  
C:Accession: J80597  
R:Kretzschmar, J.; Haendler, B.; Langer, G.; Boisdol, W.; Bringmann, P.; Alagon, A.; Don  
Gene 105, 229-237, 1991  
A:Title: The plasminogen activator family from the salivary gland of the vampire bat Des  
A:Reference number: J80597; MUID:92039036; PMID:1937019  
A:Accession: J80597  
A:Molecule type: mRNA  
A:Residues: 1-477 <KRA>  
A:Cross-references: GB:M63987; NID:q166070; PIDN:AAA31591.1; PID:g166071  
C:Superfamily: tissue plasminogen activator; EGF homology; fibronectin type I repeat hom  
C:Keywords: fibrinolysis; glycoprotein; hydrolase; kringle; serine proteinase  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-36/Domain: propeptide #status predicted <PRO>  
F:37-477/Product: plasminogen activator alpha-1 #status predicted <PLA>  
F:42-79/Domain: fibronectin type I repeat homology <1FA>  
F:87-120/Domain: EGF homology <EGF>  
F:128-209/Domain: kringle homology <KRG>  
F:226-471/Domain: trypsin homology <TRY>  
F:42-72,70-79,87-98,92-103,111-120,128-209,149-191,180-204,214-345,257-273,265-334,359-4  
F:153,398/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:225-226/Cleavage site: His-Ser (plasmin) #status predicted  
F:272,321,428/Active site: His, Asp, Ser #status predicted  
Query Match 22.3%; Score 704.5; DB 2; Length 477;  
Best Local Similarity 35.9%; Pred. No. 1.5e-36;  
Matches 161; Conservative 64; Mismatches 160; Indels 63; Gaps 15;  
QY 120 CGRGCLITQSPPYRVCVKCHPTGSCSQVVPV--CRPNPCQNGATCSRHKRSKFTCA 177  
DB 72 CDRGQA-----RC-----HTVPVNSCGEPFCFNGTQWQAVYPSDFVCQ 110  
QY 178 CPDPKGFCEIGS-DDCIVGDSYRGKQNTVQHACLWNSHLLQENYFMEDAE 236  
DB 111 CPAGYTGKRCVDTRATCYEGQGVYRGTWSTAESRVEGINWNSLLTTRTYNGRMPDAF 170  
QY 237 THGIGHNFCRNPDADEKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEESPTETKLP 296  
DB 171 NLGLGNHYCRNPNGAPKWCVV--IKAGFTSESSVPCSKA----- 212  
QY 297 GFDSCGKTEIARKIKRIYGGPKSTAGKHPWQAS-QSSLPFTISMPOGHFGGALIHPCW 356  
DB 213 --TCGLRKYKBPQLHST--GGLFTDITSHPWQAIFAQ--NRRSSGERFLCGGILLSCW 266  
QY 357 VLTAHHC--TDIKTRHLKVVLDGQDLKKEEFHQSFVEKIKFYSHYNERDEIPNDIAL 414  
DB 267 VJTAACCFQESYLPDQGLKVLGRTYRKVEEQTFFKVKYIVHKEFD--DDTYNDIAL 324  
QY 415 LKLKPDVGHCALESKYVTKVLCIPDGS--FPSSGECHISGWVETGKG--SSQLLDKVK 470  
DB 325 LQLKSDSPQCAQESDSVAICLPEANLQLPDWTCELSGYGKHSSPPYSQJKEGHR 384  
QY 471 LTANTLCNRSQLYDHMDISMTICAGNLQK-----PG-QDPCCGDSGGLTCEKDGYYVYG 525  
DB 385 LYPSSRCAPKFLFNKVTITNMMLCAGDTRSGEIVPNVHDCQGSQGGPLVCMNDNENTLLG 444

QY 526 IVSWGLECGKR--PGVYTVQVTKFLNWK 551  
DB 445 IISWGVCGEKDVPGVYTKVTINYLGIWR 472

## RESULT 14

S18932

u-plasminogen activator (EC 3.4.21.73) precursor - rat  
N:Alternate names: plasminogen activator, urokinase-type; urinary plasminogen activator  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 18-Oct-1989 #sequence\_revision 10-Feb-1995 #text\_change 18-Jun-1999  
C:Accession: S24604; I60186; I53472; S18932  
R:Rabbani, S.A.  
submitted to the EMBL Data Library, April 1992  
A:Reference number: S24604  
A:Accession: S24604

A:Molecule type: mRNA  
A:Residues: 1-15, 'H', 17-23, 'G', 25-331, 'N', 333-432 <RAB>  
A:Cross-references: EMBL:X65651; NID:g57456; PIDN:CAA46601.1; PID:g57457  
R:Experimental source: tissue kidney  
R:Handerson, B.R.; Tansey, W.P.; Phillips, S.M.; Ramshaw, I.A.; Kefford, R.F.  
Cancer Res. 52, 2489-2496, 1992  
A:Title: Transcriptional and posttranscriptional activation of urokinase plasminogen act  
A:Reference number: I60186; MUID:92233409; PMID:1568219  
A:Accession: I60186  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-432 <RES>  
A:Cross-references: EMBL:X63434; NID:g57465; PIDN:CAA45028.1; PID:g57466  
R:Experimental source: strain Fischer 344; tissue mammary  
R:Ragno, P.; Cassano, S.; Degen, J.; Kessler, C.; Blasi, F.; Rossi, G.  
FEBS Lett. 306, 193-198, 1992  
A:Title: The receptor for the plasminogen activator of urokinase type is up-regulated in  
A:Reference number: I53472; MUID:92339549; PMID:1321734  
A:Accession: I53472  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 31-62 <RE2>  
A:Cross-references: EMBL:X66907; NID:g396200; PIDN:CAA47356.1; PID:g938279  
C:Genetics:  
A:Gene: uPA  
C:Superfamily: urokinase-type plasminogen activator; EGF homology; kringle homology; try  
C:Keywords: glycoprotein; heterodimer; hydrolase; kringle; serine proteinase  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-177/Product: urokinase-type plasminogen activator chain A #status predicted <ACH>  
F:31-62/Domain: EGF homology <EGF>  
F:70-151/Domain: kringle homology <KRG>  
F:179-432/Product: urokinase-type plasminogen activator chain B #status predicted <BCH>  
F:179-420/Domain: trypsin homology <TRY>  
F:158-300,210-226,215-289,314-383,346-362,373-401/Diulfide bonds: #status predicted  
F:225,276,377/Active site: His, Asp, Ser #status predicted  
Query Match 22.3%; Score 704; DB 1; Length 432;  
Best Local Similarity 37.9%; Pred. No. 1.5e-36;  
Matches 157; Conservative 64; Mismatches 163; Indels 30; Gaps 14;

QY 159 CQNGATCSRHKRSKFTCA-CPDPKGFCEIG-SDDCYVGDGYSYRGKQNTVQHACL 216  
DB 33 CQNGCVCTRYKYSSTRSCPKPKFKGECEIDTSTKYHGNGQSYRGKANTDKGRCL 92  
QY 217 YWNSHLLQENYFMEDAEETHGIBHFCRNPDADEKPCWCFIKVTNDKVKWEYCDVSAC 276  
DB 93 AwnSPAVLQQTYNAAHRSALSLGLGKHNYCRPNQRRPWCYVQIGLKQF-----VQEC 146  
QY 277 SAQVAVPEESPTETKLP--PGFDSGKTEIARKIKRIYGGPKSTAGKHPWQASLQSS 334  
DB 147 MVQDCSLSK----KPSSTVDDQQGF--CQQLALRPR--FKLVGGEFTVVENQPFALY-- 197  
QY 335 LPLTISMPOGHFGGALIHPCWVLTAAHC--TDIKTRHLKVVLDGQDLKKEEFHQSFV 392  
DB 198 LKNGGSPSPKCGGSLISPCWASATHCFVNPQKKEEYVYVYLGQSKRNSYNPGEMKFEV 257



```

OM protein - protein search, using sw model

Run on:      May 24, 2004, 09:40:10 ; Search time 11.5 Seconds
              (without alignments)
              2535.590 Million cell updates/sec

```

Sequence: 1 MFARMSDLHVLLMALVGKT.....TQVTKFLNWKATIKSESGF 560

```

Gapop 10.0 , Gapext 0.5

Searched:      141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :      SwissProt_42.*

```

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	832.5	26.4	655	1	HGPA_HUMAN	Q04756 homo sapien
2	829	26.3	653	1	HGPA_MOUSE	Q0r098 mus musculus
3	773.5	24.5	603	1	FA12_CAVPO	Q04962 cavia porce
4	766.5	24.3	582	1	TPA_HUMAN	P00750 homo sapien
5	757.5	24.0	559	1	TPA_RAT	P19637 rattus norv
6	753.5	23.9	559	1	TPA_MOUSE	P12124 mus musculus
7	741	23.5	615	1	FA12_HUMAN	P00748 homo sapien
8	734.5	23.3	566	1	TPA_BOVIN	Q28198 bos taurus
9	731	23.2	477	1	TPA2_DESRO	P15638 desmodus ro
10	720.5	22.8	431	1	URTB_DESRO	P98121 desmodus ro
11	706.5	22.4	533	1	FA12_BOVIN	P98140 bos taurus
12	704.5	22.3	477	1	URTL_DESRO	P98119 desmodus ro
13	704	22.3	432	1	UROK_RAT	P29598 rattus norv
14	686	21.8	442	1	UROK_PIG	P04185 sus scrofa
15	683.5	21.7	433	1	UROK_BOVIN	Q05889 bos taurus
16	680	21.6	433	1	UROK_MOUSE	Q05869 mus musculus
17	660	20.9	433	1	UROK_PAPCY	P16227 papio cynoc
18	660	20.9	812	1	PLNM_BOVIN	P05868 bos taurus
19	658.5	20.9	431	1	UROK_HUMAN	P00749 homo sapien
20	650.5	20.6	434	1	UROK_CHICK	P15120 gallus gall
21	643.5	20.4	394	1	URTG_DESRO	P49150 desmodus ro
22	642.5	20.4	810	1	PLNM_HUMAN	P00747 homo sapien
23	635.5	20.1	810	1	PLNM_MACMU	P12545 macaca mula
24	607	19.2	790	1	PLNM_PIG	P05867 sus scrofa
25	603.5	19.1	812	1	PLNM_MOUSE	P28918 mus musculus
26	592.5	18.8	4548	1	APOA_HUMAN	P09519 homo sapien
27	578.5	18.3	1420	1	APOA_MACMU	P14417 macaca mula
28	571.5	18.1	810	1	PLNM_ERIEU	Q29485 erinaceus e
29	558	17.7	338	1	PLNM_HORSE	P80010 equus cabal
30	545	17.3	333	1	PLNM_CANEA	P80009 canis famil
31	535.5	17.0	343	3	PLNM_SHEEP	P80010 equus cabal
32	467.5	14.8	475	1	FA10_CHICK	P81286 ovnis aries
33	462.5	14.7	761	1	NETR_MOUSE	P25155 gallus gall
						O08762 mus musculus

[illegible]

RT "Mouse hepatocyte growth factor activator."  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-2126753; PubMed-11032833;  
 RA van Adelsberg J.S., Sehgal S., Kukes A., Brady C., Barasch J.,  
 Yang J., Huan Y.;  
 RT "Activation of hepatocyte growth factor (HGF) by endogenous HGF  
 activator is required for metanephric kidney morphogenesis in  
 vitro".  
 RT J. Biol. Chem. 276:15099-15106(2001).  
 CC -!- FUNCTION: Activates hepatocyte growth factor (HGF) by converting  
 it from a single chain to a heterodimeric form (By similarity).  
 CC -!- SUBUNIT: Dimer of a short chain and a long chain linked by a  
 disulfide bond (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted as an inactive single-chain  
 precursor and is then activated to a heterodimeric form (By  
 similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family S1.  
 CC -!- SIMILARITY: Contains 2 EGF-like domains.  
 CC -!- SIMILARITY: Contains 1 fibronectin type I domain.  
 CC -!- SIMILARITY: Contains 1 fibronectin type II domain.  
 CC -!- SIMILARITY: Contains 1 kringle domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF099017; AAF02489.1; -;  
 DR EMBL; AF224724; AAF34712.1; -;  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.228; -;  
 DR MGD; MGI:1859281; Hgfac.  
 DR InterPro; IPR009003; Cys\_ser\_trypsin.  
 DR InterPro; IPR007422; EGF\_2.  
 DR InterPro; IPR006209; EGF\_Like.  
 DR InterPro; IPR000083; Fibrinctnl.  
 DR InterPro; IPR000562; FN\_Type\_II.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR000001; Kringle.  
 DR InterPro; IPR001254; Peptidase\_S1A.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR Pfam; PF00008; EGF; 2.  
 DR Pfam; PF00039; fn1; 1.  
 DR Pfam; PF00040; fn2; 1.  
 DR Pfam; PF00051; kringle; 1.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00013; ENTYPDI.  
 DR PRINTS; PR00018; KRINGLE.  
 DR ProDom; PD000995; FN\_Type\_II; 1.  
 DR ProDom; PD000395; Kringle; 1.  
 DR SMART; SM00181; EGF; 2.  
 DR SMART; SM00059; FN2; 1.  
 DR SMART; SM00130; KR; 1.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00026; EGF\_3; 2.  
 DR PROSITE; PS01253; FIBRONECTIN\_1; 1.  
 DR PROSITE; PS00023; FIBRONECTIN\_2; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 1.  
 DR PROSITE; PS00020; KRINGLE\_2; 1.  
 DR PROSITE; PS02470; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Glycoprotein; Plasma; Serine protease; Kringle; Signal;  
 KW EGF-like domain; Repeat; Zymogen.  
 FT SIGNAL 1 29 BY SIMILARITY.

FT PROPEP 30 369 CLEAVED IN ACTIVE FORM (BY SIMILARITY).  
 FT CHAIN 370 405 HEPATOCYTE GROWTH FACTOR ACTIVATOR SHORT  
 FT CHAIN 406 653 HEPATOCYTE GROWTH FACTOR ACTIVATOR LONG  
 FT CHAIN 406 653 CHAIN.  
 FT DOMAIN 105 145 FIBRONECTIN TYPE-II.  
 FT DOMAIN 157 195 EGF-LIKE 1.  
 FT DOMAIN 197 237 FIBRONECTIN TYPE-I.  
 FT DOMAIN 238 276 EGF-LIKE 2.  
 FT DOMAIN 283 364 KRINGLE.  
 FT DOMAIN 406 653 SERINE PROTEASE.  
 FT ACT\_SITE 445 445 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 495 495 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 596 596 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 105 130 BY SIMILARITY.  
 FT DISULFID 119 145 BY SIMILARITY.  
 FT DISULFID 161 172 BY SIMILARITY.  
 FT DISULFID 166 183 BY SIMILARITY.  
 FT DISULFID 185 194 BY SIMILARITY.  
 FT DISULFID 199 227 BY SIMILARITY.  
 FT DISULFID 225 234 BY SIMILARITY.  
 FT DISULFID 242 253 BY SIMILARITY.  
 FT DISULFID 247 264 BY SIMILARITY.  
 FT DISULFID 266 275 BY SIMILARITY.  
 FT DISULFID 283 364 BY SIMILARITY.  
 FT DISULFID 304 346 BY SIMILARITY.  
 FT DISULFID 335 359 BY SIMILARITY.  
 FT DISULFID 392 519 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 430 446 BY SIMILARITY.  
 FT DISULFID 438 508 BY SIMILARITY.  
 FT DISULFID 533 602 BY SIMILARITY.  
 FT DISULFID 565 581 BY SIMILARITY.  
 FT DISULFID 592 620 BY SIMILARITY.  
 FT CARBOHYD 39 39 N-LINKED (GLCNAC. ) (POTENTIAL).  
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. ) (POTENTIAL).  
 FT CARBOHYD 63 63 N-LINKED (GLCNAC. ) (POTENTIAL).  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. ) (POTENTIAL).  
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. ) (POTENTIAL).  
 FT CARBOHYD 544 544 N-LINKED (GLCNAC. ) (POTENTIAL).  
 FT CONFLICT 164 164 G -> W (IN REF. 2).  
 SQ SEQUENCE 653 AA; 70567 MW; 88B4B20255DF7FDC CRC64;  
 Query Match 26.3%; Score 829; DB 1; Length 653;  
 Best Local Similarity 34.4%; Pred. No. 4.4e-51;  
 Matches 190; Conservative 79; Mismatches 187; Indels 96; Gaps 22;  
 Qy 60 THAENPD--WYTEDQA-----DPCQNPCEHGGDCLV---HGSTFTCSCLAFFSG 105  
 Db 133 THNYDRDRAWGYCAEVTLPVEGPAILDPCASGCLNGGTCSSTHDSGS-YHSCSCLAFFTG 191  
 Qy 106 NKQKQVQNTCKD-----NPGCRGQCLITQSPYRYRCVKCHPYTG 144  
 Db 192 KDCG--TEKCFDTRRYEYFEVGDHARVSEGHVEQCG---CMGQA---RCEDTHH--- 239  
 Qy 145 PSCSQVVPVCRPNPCONGATCSRHKRSKFTACPDQFQKFCFCEI-GSDDCYVDGYSYR 203  
 Db 240 -----TACLSSPCLNGGTCHLIVGTGTSVCTCPLGYAGRCFNIVPTHECFLNGTEYR 292  
 Qy 204 GKMRRTVQHACLYWNSHLLLOENWFMEDASTHGIGEHNFNCPNPDADKPCWCFIKVTN 263  
 Db 293 GVASTAASGLSCLAWNSDLLYQELHVDVSAVAALLGLGPHAYCRNFDKDEPWCYV-VKD 351  
 Qy 264 DKVWEYCDVSACSADQ-----VAYPESPTERTKLPGFDSQKTEIAERKIK- 312  
 Db 352 NALSWEYCRUTACESLARVHSQTPETLAALPESAPAVRPT-----CGKRHKRKTFLRP 404  
 Qy 313 RIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCGGLIHPCWLVLTAAHC--TDIKTRH 370  
 Db 405 RIIGSSSLPGSHPLAAIYG-----NSFCAGSLVHTCWVSAHCFANSPPRDS 455  
 Qy 371 LKVLGQDQDLKKEEFHQSFVEKIFKYSHYNERDEIPHN-DIALKLKLPVDGHCALSK 429  
 Db 456 ITVLGQHFFNRRTDVTQTGIEKYVPYTLYSVEN--PNNHDLVLRLLKKKGRCARVSRQ 513





QY 70 ---TBDQADPC-QPNPCBHGDCLVHGSTFTSCSLAPFSGNKKQKVQNTCKDNPCGRGQC 125  
DB 87 CLEPKVKVDHCKSNPCQGGGICVNTLSSPHCLCPDHLTGKHCQ-----REKC 134  
QY 126 LITQSPPPY-----RCVCKHPVTGSCSQV-VPVCRPNPCONGATCSRHK 169  
DB 135 FEPQHRFFHENEIWFRTGPGAVAKCHKGP--DAHCKQMSQBCQTNPCINGRCL--E 190  
QY 170 RESKTCACPDQFKGKCEIGSD-DCYVGDGYSYRGKVRNVQNHACLYWNHLLLOBNY 228  
DB 191 VEGHLLDCPMGYTGPFCDDLTATASCYEGRGVSYRGMAITTVSGAKCQRMASEATYR--- 247  
QY 229 NMFEMDAETHGGEHNFRCNPDADKPKCFIKVTNDKVKWECDVSACS--AODVAYPE- 285  
DB 248 NNTAQALRRLGHHTFCRNPNDTRPWCFFVMGN-RLSWEYCDLQAQCOYPPPTATPHD 306  
QY 285 --ESPTSTK-----LFGFDS--CGKTEAER--KIKRIYGGPKS 320  
DB 307 REPHKPLSSRLSILQTPQTTQNALANELPETSSLLCGQ-RLKRLSSLSRIVGGIVA 365  
QY 321 TAGKHPWQASLOSSIPLTISMQGHFCGGLIHPCWVLTAAHCTDIK--TBHLKXVLGDQ 378  
DB 366 LFGAHPYTAALYWG-----SNFCSGLIAPCWLTAAHCLQNRPAPELKVVLGDQ 416  
QY 379 DLKKEEFHEQ-----SFRVEKIFKYSYNERDEIPHNDIALKL-KPVDGHCALESKYVK 432  
DB 417 RHNSQCEHCQTLAVHSYRLHEAFSPSY-----LNDLALLRLQXSADGSCAQLSPVYQ 469  
QY 433 TVCLPDG-SFPGSE---CHISGWGVTEG--KGSRLQDADKVLANTLNSRQLYDHM 486  
DB 470 TVCLPSGAPPSESETTCCEVAGNQHPEGAEYSFLQEAQVPLISSERSSPEVHGDA 529  
QY 487 IDDSMICAGNLCKPGQDTCQSGGGLTCCKDGTYY---VVGIVSWGLECKGR--PGVYT 541  
DB 530 FLISGMLCAGFLR-GTDCACQSGGGLVCEDEAAEHRILLRIGVSWGSGCGDRNKPQVYT 588  
QY 542 QVTKFLNWK 551  
DB 589 DVASYLTWIQ 598

## RESULT 4

ID TPA\_HUMAN STANDARD; PRT; 562 AA.  
AC P00750; Q15103;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)  
GN PLAT.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]\_TaxID=9606;  
RP SEQUENCE FROM N.A.  
RC TISSUE=Melanoma;  
RX MEDLINE=83115262; PubMed=6337343;  
RA Pennica D., Holmes W.E., Kohr W.J., Harkins R.N., Vehar G.A.,  
RA Ward C.A., Bennett W.F., Yelverton E., Seeburg P.H., Heyneker H.L.,  
RA Goeddel D.V., Collen D.;  
RT "Cloning and expression of human tissue-type plasminogen activator  
cDNA in E. coli.";  
RL Nature 301:214-221 (1983).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal lung;  
RX MEDLINE=88262579; PubMed=3133640;  
RA Sasaki H., Saico Y., Hayashi M., Otsuka K., Niwa M.;  
RT "Nucleotide sequence of the tissue-type plasminogen activator cDNA  
from human fetal lung cells.";

RL Nucleic Acids Res. 16:5695-5695 (1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88054470; PubMed=2824147;  
RA Reddy V.B., Garrazone A.J., Sasek H., Wei C.-M., Watkins P., Galli J.,  
RA Hsiung N.;  
RT "Expression of human uterine tissue-type plasminogen activator in  
mouse cells using BPV vectors.";  
RL DNA 6:461-472 (1987).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86196143; PubMed=3009482;  
RA Friezner Degen S.J., Rajput B., Reich E.;  
RT "The human tissue plasminogen activator gene.";  
RL J. Biol. Chem. 261:6972-6985 (1986).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84298137; PubMed=6089198;  
RA NY T., Elgh F., Lund B.;  
RT "The structure of the human tissue-type plasminogen activator gene:  
correlation of intron and exon structures to functional and  
structural domains.";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:5355-5359 (1984).  
RN [6]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86284200; PubMed=3090401;  
RA Harris T.J., Patel T., Marston P.A., Little S., Emtage J.S.,  
RA Odenakker G., Volckaert G., Rombauts W., Billiau A., Somer P.;  
RT "Cloning of cDNA coding for human tissue-type plasminogen activator  
and its expression in Escherichia coli.";  
RL Mol. Biol. Med. 3:279-292 (1986).  
RN [7]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
RC TISSUE=Umbilical vein;  
RX MEDLINE=90192129; PubMed=2107528;  
RA Siebert P.D., Fong K.;  
RT "Variant tissue-type plasminogen activator (PLAT) cDNA obtained from  
human endothelial cells.";  
RL Nucleic Acids Res. 18:1086-1086 (1990).  
RN [8]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=22389257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [9]  
RP SEQUENCE OF 212-361 FROM N.A.  
RX MEDLINE=83169656; PubMed=6572897;  
RA Edlund T., Ny T., Raanby M., Heden L.-O., Palm G., Holmgren E.,  
RA Josephson S.;  
RT "Isolation of cDNA sequences coding for a part of human tissue  
plasminogen activator.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:349-352 (1983).  
RN [10]  
RP SEQUENCE OF 1-36 FROM N.A.

RX MEDLINE=85289338; PubMed=3161893;  
 RA Fisher R., Waller E.K., Grossi G., Thompson D., Tizard R.,  
 RA Schleuning W.-D.;  
 RT "Isolation and characterization of the human tissue-type plasminogen  
 RT activator structural gene including its 5' flanking region.";  
 RL J. Biol. Chem. 260:11223-11230(1985).  
 RN [11]  
 RN SEQUENCE OF 31-562 FROM N.A.  
 RP MEDLINE=91291340; PubMed=1369681;  
 RA Itagaki Y., Yasuda H., Morinaga T., Mitsuda S., Higashio K.;  
 RA "Purification and characterization of tissue plasminogen activator  
 RT secreted by human embryonic lung diploid fibroblasts, IMR-90 cells.";  
 RL Agric. Biol. Chem. 55:1225-1232(1991).  
 RN [12]  
 RN SEQUENCE OF 36-562.  
 RP TISSUE=Melanoma;  
 RC MEDLINE=85000468; PubMed=6433976;  
 RA Pohl G., Kaelinroem M., Bergsdorf N., Wallen P., Joernvall H.;  
 RT "Tissue plasminogen activator: peptide analyses confirm an indirectly  
 RT derived amino acid sequence, identify the active site serine residue,  
 RT establish glycosylation sites, and localize variant differences.";  
 RL Biochemistry 23:3701-3707(1984).  
 RN [13]  
 RN SEQUENCE OF 33-52 AND 311-330.  
 RP TISSUE=Melanoma;  
 RC MEDLINE=83209620; PubMed=6682760;  
 RA Wallen P., Pohl G., Bergsdorf N., Raanby M., Ny T., Joernvall H.;  
 RT "Purification and characterization of a melanoma cell plasminogen  
 RT activator.";  
 RL Eur. J. Biochem. 132:681-686(1983).  
 RN [14]  
 RN STRUCTURE OF CARBOHYDRATES.  
 RP MEDLINE=90092112; PubMed=2513186;  
 RA Pfeiffer G., Schmidt M., Strube K.-H., Geyer R.;  
 RT "Carbohydrate structure of recombinant human uterine tissue  
 RT plasminogen activator expressed in mouse epithelial cells.";  
 RL Eur. J. Biochem. 186:273-286(1989).  
 RN [15]  
 RN CARBOHYDRATE-LINKAGE SITE THR-96.  
 RP MEDLINE=91159408; PubMed=1900431;  
 RA Harris R.J., Leonard C.K., Guzzetta A.W., Spellman M.W.;  
 RT "Tissue plasminogen activator has an O-linked fucose attached to  
 RT threonine-61 in the epidermal growth factor domain.";  
 RL Biochemistry 30:2311-2314(1991).  
 RN [16]  
 RN DISULFIDE BONDS IN KRINGLE 2.  
 RP MEDLINE=91244765; PubMed=1645336;  
 RA Vlahos C.J., Wilhelm O.G., Hassell T., Jaskunas S.R., Bang N.U.;  
 RT "Disulfide pairing of the recombinant kringle-2 domain of tissue  
 RT plasminogen activator produced in Escherichia coli.";  
 RL J. Biol. Chem. 266:10070-10072(1991).  
 RN [17]  
 RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF CATALYTIC DOMAIN.  
 RP MEDLINE=96200985; PubMed=8613982;  
 RA Lamba D., Bauer M., Huber R., Fischer S., Rudolph R., Kohnert U.,  
 RA Bode W.;  
 RT "The 2.3 A crystal structure of the catalytic domain of recombinant  
 RT two-chain human tissue-type plasminogen activator.";  
 RL J. Mol. Biol. 258:117-135(1996).  
 RN [18]  
 RN X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF CATALYTIC DOMAIN.  
 RP MEDLINE=97449126; PubMed=9305622;  
 RA Renatus M., Engh R.A., Stubbs M.T., Huber R., Fischer S., Kohnert U.,  
 RA Bode W.;  
 RT "Lysine 156 promotes the anomalous proenzyme activity of tPA: X-ray  
 RT crystal structure of single-chain human tPA.";  
 RL EMBO J. 16:4797-4805(1997).  
 RN [19]  
 RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF KRINGLE 2.  
 RP MEDLINE=92118803; PubMed=1310033;  
 RA de Vos A., Ultsch M.H., Kelley R.F., Padmanabhan K., Tulinsky A.,  
 RA Westbrook M.L., Kosiakof A.A.;  
 RT "Crystal structure of the kringle 2 domain of tissue plasminogen

RT activator at 2.4-A resolution.";  
 RL Biochemistry 31:270-279(1992).  
 RN [20]  
 RP STRUCTURE BY NMR OF KRINGLE 2.  
 RX MEDLINE=90122799; PubMed=2558718;  
 RA Byeon I.-J.L., Kelley R.F., Llinas M.;  
 RT "1H NMR structural characterization of a recombinant kringle 2 domain  
 RT from human tissue-type plasminogen activator.";  
 RL Biochemistry 28:9350-9360(1989).  
 RN [21]  
 RP STRUCTURE BY NMR OF KRINGLE 2.  
 RX MEDLINE=91200042; PubMed=1901789;  
 RA Byeon I.-J.L., Kelley R.F., Llinas M.;  
 RT "Kringle-2 domain of the tissue-type plasminogen activator. 1H-NMR  
 RT assignments and secondary structure.";  
 RL Eur. J. Biochem. 197:155-165(1991).  
 RN [22]  
 RP STRUCTURE BY NMR OF KRINGLE 2.  
 RX MEDLINE=92106329; PubMed=1762144;  
 RA Byeon I.-J.L., Llinas M.;  
 RT "Solution structure of the tissue-type plasminogen activator kringle  
 RT 2 domain complexed to 6-aminohexanoic acid an antifibrinolytic  
 RT drug.";  
 RL J. Mol. Biol. 222:1035-1051(1991).  
 RN [23]  
 Query Match 24.3%; Score 766.5; DB 1; Length 562;  
 Best Local Similarity 35.0%; Pred. No. 9.4e-47;  
 Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;  
 QY 77 QCPNPGHGGDC--LVHSGTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG----- 123  
 DB 86 CSEPRCFNGGTCQALYFDFVCCPEGFAGKCEIDTRATCYEDQGISYRGWSTAESG 145  
 QY 124 -QCLITQSPFYRCVCKHPYTPGSCSVVPCVRNPQ-----NGATCSRHKRRKFTAC 178  
 DB 146 AECTNWS-----SALAQKPSGR-----RPDAIRLGIGNHNYCRNPDSDKPCYV 192  
 QY 179 PDQFK-GKF-----CEIGSDDCYVGQSVYRGKMRVNOHACLYWNSHLLLOENY 228  
 DB 193 ---FRAGKYSSEFCSTPACSEGNDCYFGNSAARGTHSLTESGASCLPWNMILICKVY 249  
 QY 229 NMFMDAETHGIGEHNFGRNPDADEKPCFIKVTNDKYKWEYCDVSACSAQDVAYPEESP 288  
 DB 250 TQNPESAQALGLGKXNYCRNPDGDAKPCWCHV-LKNRLTWECYCDVPSCS----- 297  
 QY 289 TEPSTKLPFGFSCGKTEAERKIKRIYGGFKSTAGKHPQASLQSLPLTITSMQGHFCG 348  
 DB 298 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAAIFAK--HRRSPGRFLCG 343  
 QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVVLGQDLKKEEFHEQSPRVEKIPKYSYNERDE 406  
 DB 344 GLIISCSWILSAAHCFQERFPFHLTVLGTIVYVPGEEQKFEVEKYIVHKEFD--DD 401  
 QY 407 IPHNDIALLKPFVDGHCALSKYKVKVCLP--DGSPPSGSECHISGHWVTETGKG--SR 462  
 DB 402 TYDNDIALQLKSDSSRCAQESSVVRVTCPLPPADLPQPDWTECELSGYGKHEALSPFYSE 461  
 QY 463 QLLDAKVKLIANTLCLNSRQLYDMDSDMICAGNLQKPG-----QDTCCGSGGGLTCEK 517  
 DB 462 RUKERAHVRLYPSRRTSGLNLRVTDNMLCAGTRSGGPGQANLHDAQCQSGDGLVCLN 521  
 QY 518 DGTYYVYGVISWGLECGKR--PGVYTVQTKFLNLIKATIK 555  
 DB 522 DGRMTLVGIISWGLGCGKQVGVYTKVTNYLDWIRDNR 561  
 RESULT 5  
 ID TPA RAT  
 AC PI9637;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)  
 DE (t-PA) (t-plasminogen activator).  
 GN PLAT.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RN [2]  
 RN [3]  
 RN [4]  
 RN [5]  
 RN [6]  
 RN [7]  
 RN [8]  
 RN [9]  
 RN [10]  
 RN [11]  
 RN [12]  
 RN [13]  
 RN [14]  
 RN [15]  
 RN [16]  
 RN [17]  
 RN [18]  
 RN [19]  
 RN [20]  
 RN [21]  
 RN [22]  
 RN [23]  
 RN [24]  
 RN [25]  
 RN [26]  
 RN [27]  
 RN [28]  
 RN [29]  
 RN [30]  
 RN [31]  
 RN [32]  
 RN [33]  
 RN [34]  
 RN [35]  
 RN [36]  
 RN [37]  
 RN [38]  
 RN [39]  
 RN [40]  
 RN [41]  
 RN [42]  
 RN [43]  
 RN [44]  
 RN [45]  
 RN [46]  
 RN [47]  
 RN [48]  
 RN [49]  
 RN [50]  
 RN [51]  
 RN [52]  
 RN [53]  
 RN [54]  
 RN [55]  
 RN [56]  
 RN [57]  
 RN [58]  
 RN [59]  
 RN [60]  
 RN [61]  
 RN [62]  
 RN [63]  
 RN [64]  
 RN [65]  
 RN [66]  
 RN [67]  
 RN [68]  
 RN [69]  
 RN [70]  
 RN [71]  
 RN [72]  
 RN [73]  
 RN [74]  
 RN [75]  
 RN [76]  
 RN [77]  
 RN [78]  
 RN [79]  
 RN [80]  
 RN [81]  
 RN [82]  
 RN [83]  
 RN [84]  
 RN [85]  
 RN [86]  
 RN [87]  
 RN [88]  
 RN [89]  
 RN [90]  
 RN [91]  
 RN [92]  
 RN [93]  
 RN [94]  
 RN [95]  
 RN [96]  
 RN [97]  
 RN [98]  
 RN [99]  
 RN [100]  
 RN [101]  
 RN [102]  
 RN [103]  
 RN [104]  
 RN [105]  
 RN [106]  
 RN [107]  
 RN [108]  
 RN [109]  
 RN [110]  
 RN [111]  
 RN [112]  
 RN [113]  
 RN [114]  
 RN [115]  
 RN [116]  
 RN [117]  
 RN [118]  
 RN [119]  
 RN [120]  
 RN [121]  
 RN [122]  
 RN [123]  
 RN [124]  
 RN [125]  
 RN [126]  
 RN [127]  
 RN [128]  
 RN [129]  
 RN [130]  
 RN [131]  
 RN [132]  
 RN [133]  
 RN [134]  
 RN [135]  
 RN [136]  
 RN [137]  
 RN [138]  
 RN [139]  
 RN [140]  
 RN [141]  
 RN [142]  
 RN [143]  
 RN [144]  
 RN [145]  
 RN [146]  
 RN [147]  
 RN [148]  
 RN [149]  
 RN [150]  
 RN [151]  
 RN [152]  
 RN [153]  
 RN [154]  
 RN [155]  
 RN [156]  
 RN [157]  
 RN [158]  
 RN [159]  
 RN [160]  
 RN [161]  
 RN [162]  
 RN [163]  
 RN [164]  
 RN [165]  
 RN [166]  
 RN [167]  
 RN [168]  
 RN [169]  
 RN [170]  
 RN [171]  
 RN [172]  
 RN [173]  
 RN [174]  
 RN [175]  
 RN [176]  
 RN [177]  
 RN [178]  
 RN [179]  
 RN [180]  
 RN [181]  
 RN [182]  
 RN [183]  
 RN [184]  
 RN [185]  
 RN [186]  
 RN [187]  
 RN [188]  
 RN [189]  
 RN [190]  
 RN [191]  
 RN [192]  
 RN [193]  
 RN [194]  
 RN [195]  
 RN [196]  
 RN [197]  
 RN [198]  
 RN [199]  
 RN [200]  
 RN [201]  
 RN [202]  
 RN [203]  
 RN [204]  
 RN [205]  
 RN [206]  
 RN [207]  
 RN [208]  
 RN [209]  
 RN [210]  
 RN [211]  
 RN [212]  
 RN [213]  
 RN [214]  
 RN [215]  
 RN [216]  
 RN [217]  
 RN [218]  
 RN [219]  
 RN [220]  
 RN [221]  
 RN [222]  
 RN [223]  
 RN [224]  
 RN [225]  
 RN [226]  
 RN [227]  
 RN [228]  
 RN [229]  
 RN [230]  
 RN [231]  
 RN [232]  
 RN [233]  
 RN [234]  
 RN [235]  
 RN [236]  
 RN [237]  
 RN [238]  
 RN [239]  
 RN [240]  
 RN [241]  
 RN [242]  
 RN [243]  
 RN [244]  
 RN [245]  
 RN [246]  
 RN [247]  
 RN [248]  
 RN [249]  
 RN [250]  
 RN [251]  
 RN [252]  
 RN [253]  
 RN [254]  
 RN [255]  
 RN [256]  
 RN [257]  
 RN [258]  
 RN [259]  
 RN [260]  
 RN [261]  
 RN [262]  
 RN [263]  
 RN [264]  
 RN [265]  
 RN [266]  
 RN [267]  
 RN [268]  
 RN [269]  
 RN [270]  
 RN [271]  
 RN [272]  
 RN [273]  
 RN [274]  
 RN [275]  
 RN [276]  
 RN [277]  
 RN [278]  
 RN [279]  
 RN [280]  
 RN [281]  
 RN [282]  
 RN [283]  
 RN [284]  
 RN [285]  
 RN [286]  
 RN [287]  
 RN [288]  
 RN [289]  
 RN [290]  
 RN [291]  
 RN [292]  
 RN [293]  
 RN [294]  
 RN [295]  
 RN [296]  
 RN [297]  
 RN [298]  
 RN [299]  
 RN [300]  
 RN [301]  
 RN [302]  
 RN [303]  
 RN [304]  
 RN [305]  
 RN [306]  
 RN [307]  
 RN [308]  
 RN [309]  
 RN [310]  
 RN [311]  
 RN [312]  
 RN [313]  
 RN [314]  
 RN [315]  
 RN [316]  
 RN [317]  
 RN [318]  
 RN [319]  
 RN [320]  
 RN [321]  
 RN [322]  
 RN [323]  
 RN [324]  
 RN [325]  
 RN [326]  
 RN [327]  
 RN [328]  
 RN [329]  
 RN [330]  
 RN [331]  
 RN [332]  
 RN [333]  
 RN [334]  
 RN [335]  
 RN [336]  
 RN [337]  
 RN [338]  
 RN [339]  
 RN [340]  
 RN [341]  
 RN [342]  
 RN [343]  
 RN [344]  
 RN [345]  
 RN [346]  
 RN [347]  
 RN [348]  
 RN [349]  
 RN [350]  
 RN [351]  
 RN [352]  
 RN [353]  
 RN [354]  
 RN [355]  
 RN [356]  
 RN [357]  
 RN [358]  
 RN [359]  
 RN [360]  
 RN [361]  
 RN [362]  
 RN [363]  
 RN [364]  
 RN [365]  
 RN [366]  
 RN [367]  
 RN [368]  
 RN [369]  
 RN [370]  
 RN [371]  
 RN [372]  
 RN [373]  
 RN [374]  
 RN [375]  
 RN [376]  
 RN [377]  
 RN [378]  
 RN [379]  
 RN [380]  
 RN [381]  
 RN [382]  
 RN [383]  
 RN [384]  
 RN [385]  
 RN [386]  
 RN [387]  
 RN [388]  
 RN [389]  
 RN [390]  
 RN [391]  
 RN [392]  
 RN [393]  
 RN [394]  
 RN [395]  
 RN [396]  
 RN [397]  
 RN [398]  
 RN [399]  
 RN [400]  
 RN [401]  
 RN [402]  
 RN [403]  
 RN [404]  
 RN [405]  
 RN [406]  
 RN [407]  
 RN [408]  
 RN [409]  
 RN [410]  
 RN [411]  
 RN [412]  
 RN [413]  
 RN [414]  
 RN [415]  
 RN [416]  
 RN [417]  
 RN [418]  
 RN [419]  
 RN [420]  
 RN [421]  
 RN [422]  
 RN [423]  
 RN [424]  
 RN [425]  
 RN [426]  
 RN [427]  
 RN [428]  
 RN [429]  
 RN [430]  
 RN [431]  
 RN [432]  
 RN [433]  
 RN [434]  
 RN [435]  
 RN [436]  
 RN [437]  
 RN [438]  
 RN [439]  
 RN [440]  
 RN [441]  
 RN [442]  
 RN [443]  
 RN [444]  
 RN [445]  
 RN [446]  
 RN [447]  
 RN [448]  
 RN [449]  
 RN [450]  
 RN [451]  
 RN [452]  
 RN [453]  
 RN [454]  
 RN [455]  
 RN [456]  
 RN [457]  
 RN [458]  
 RN [459]  
 RN [460]  
 RN [461]  
 RN [462]  
 RN [463]  
 RN [464]  
 RN [465]  
 RN [466]  
 RN [467]  
 RN [468]  
 RN [469]  
 RN [470]  
 RN [471]  
 RN [472]  
 RN [473]  
 RN [474]  
 RN [475]  
 RN [476]  
 RN [477]  
 RN [478]  
 RN [479]  
 RN [480]  
 RN [481]  
 RN [482]  
 RN [483]  
 RN [484]  
 RN [485]  
 RN [486]  
 RN [487]  
 RN [488]  
 RN [489]  
 RN [490]  
 RN [491]  
 RN [492]  
 RN [493]  
 RN [494]  
 RN [495]  
 RN [496]  
 RN [497]  
 RN [498]  
 RN [499]  
 RN [500]  
 RN [501]  
 RN [502]  
 RN [503]  
 RN [504]  
 RN [505]  
 RN [506]  
 RN [507]  
 RN [508]  
 RN [509]  
 RN [510]  
 RN [511]  
 RN [512]  
 RN [513]  
 RN [514]  
 RN [515]  
 RN [516]  
 RN [517]  
 RN [518]  
 RN [519]  
 RN [520]  
 RN [521]  
 RN [522]  
 RN [523]  
 RN [524]  
 RN [525]  
 RN [526]  
 RN [527]  
 RN [528]  
 RN [529]  
 RN [530]  
 RN [531]  
 RN [532]  
 RN [533]  
 RN [534]  
 RN [535]  
 RN [536]  
 RN [537]  
 RN [538]  
 RN [539]  
 RN [540]  
 RN [541]  
 RN [542]  
 RN [543]  
 RN [544]  
 RN [545]  
 RN [546]  
 RN [547]  
 RN [548]  
 RN [549]  
 RN [550]  
 RN [551]  
 RN [552]  
 RN [553]  
 RN [554]  
 RN [555]  
 RN [556]  
 RN [557]  
 RN [558]  
 RN [559]  
 RN [560]  
 RN [561]  
 RN [562]  
 RN [563]  
 RN [564]  
 RN [565]  
 RN [566]  
 RN [567]  
 RN [568]  
 RN [569]  
 RN [570]  
 RN [571]  
 RN [572]  
 RN [573]  
 RN [574]  
 RN [575]  
 RN [576]  
 RN [577]  
 RN [578]  
 RN [579]  
 RN [580]  
 RN [581]  
 RN [582]  
 RN [583]  
 RN [584]  
 RN [585]  
 RN [586]  
 RN [587]  
 RN [588]  
 RN [589]  
 RN [590]  
 RN [591]  
 RN [592]  
 RN [593]  
 RN [594]  
 RN [595]  
 RN [596]  
 RN [597]  
 RN [598]  
 RN [599]  
 RN [600]  
 RN [601]  
 RN [602]  
 RN [603]  
 RN [604]  
 RN [605]  
 RN [606]  
 RN [607]  
 RN [608]  
 RN [609]  
 RN [610]  
 RN [611]  
 RN [612]  
 RN [613]  
 RN [614]  
 RN [615]  
 RN [616]  
 RN [617]  
 RN [618]  
 RN [619]  
 RN [620]  
 RN [621]  
 RN [622]  
 RN [623]  
 RN [624]  
 RN [625]  
 RN [626]  
 RN [627]  
 RN [628]  
 RN [629]  
 RN [630]  
 RN [631]  
 RN [632]  
 RN [633]  
 RN [634]  
 RN [635]  
 RN [636]  
 RN [637]  
 RN [638]  
 RN [639]  
 RN [640]  
 RN [641]  
 RN [642]  
 RN [643]  
 RN [644]  
 RN [645]  
 RN [646]  
 RN [647]  
 RN [648]  
 RN [649]  
 RN [650]  
 RN [651]  
 RN [652]  
 RN [653]  
 RN [654]  
 RN [655]  
 RN [656]  
 RN [657]  
 RN [658]  
 RN [659]  
 RN [660]  
 RN [661]  
 RN [662]  
 RN [663]  
 RN [664]  
 RN [665]  
 RN [666]  
 RN [667]  
 RN [668]  
 RN [669]  
 RN [670]  
 RN [671]  
 RN [672]  
 RN [673]  
 RN [674]  
 RN [675]  
 RN [676]  
 RN [677]  
 RN [678]  
 RN [679]  
 RN [680]  
 RN [681]  
 RN [682]  
 RN [683]  
 RN [684]  
 RN [685]  
 RN [686]  
 RN [687]  
 RN [688]  
 RN [689]  
 RN [690]  
 RN [691]  
 RN [692]  
 RN [693]  
 RN [694]  
 RN [695]  
 RN [696]  
 RN [697]  
 RN [698]  
 RN [699]  
 RN [700]  
 RN [701]  
 RN [702]  
 RN [703]  
 RN [704]  
 RN [705]  
 RN [706]  
 RN [707]  
 RN [708]  
 RN [709]  
 RN [710]  
 RN [711]  
 RN [712]  
 RN [713]  
 RN [714]  
 RN [715]  
 RN [716]  
 RN [717]  
 RN [718]  
 RN [719]  
 RN [720]  
 RN [721]  
 RN [722]  
 RN [723]  
 RN [724]  
 RN [725]  
 RN [726]  
 RN [727]  
 RN [728]  
 RN [729]  
 RN [730]  
 RN [731]  
 RN [732]  
 RN [733]  
 RN [734]  
 RN [735]  
 RN [736]  
 RN [737]  
 RN [738]  
 RN [739]  
 RN [740]  
 RN [741]  
 RN [742]  
 RN [743]  
 RN [744]  
 RN [745]  
 RN [746]  
 RN [747]  
 RN [748]  
 RN [749]  
 RN [750]  
 RN [751]  
 RN [752]  
 RN [753]  
 RN [754]  
 RN [755]  
 RN [756]  
 RN [757]  
 RN [758]  
 RN [759]  
 RN [760]  
 RN [761]  
 RN [762]  
 RN [763]  
 RN [764]  
 RN [765]  
 RN [766]  
 RN [767]  
 RN [768]  
 RN [769]  
 RN [770]  
 RN [771]  
 RN [772]  
 RN [773]  
 RN [774]  
 RN [775]  
 RN [776]  
 RN [777]  
 RN [778]  
 RN [779]  
 RN [780]  
 RN [781]  
 RN [782]  
 RN [783]  
 RN [784]  
 RN [785]  
 RN [786]  
 RN [787]  
 RN [788]  
 RN [789]  
 RN [790]  
 RN [791]  
 RN [792]  
 RN [793]  
 RN [794]  
 RN [795]  
 RN [796]  
 RN [797]  
 RN [798]  
 RN [799]  
 RN [800]  
 RN [801]  
 RN [802]  
 RN [803]  
 RN [804]  
 RN [805]  
 RN [806]  
 RN [807]  
 RN [808]  
 RN [809]  
 RN [810]  
 RN [811]  
 RN [812]  
 RN [813]  
 RN [814]  
 RN [815]  
 RN [816]  
 RN [817]  
 RN [818]  
 RN [819]  
 RN [820]  
 RN [821]  
 RN [822]  
 RN [823]  
 RN [824]  
 RN [825]  
 RN [826]  
 RN [827]  
 RN [828]  
 RN [829]  
 RN [830]  
 RN [831]  
 RN [832]  
 RN [833]  
 RN [834]  
 RN [835]  
 RN [836]  
 RN [837]  
 RN [838]  
 RN [839]  
 RN [840]  
 RN [841]  
 RN [842]  
 RN [843]  
 RN [844]  
 RN [845]  
 RN [846]  
 RN [847]  
 RN [848]  
 RN [849]  
 RN [850]  
 RN [851]  
 RN [852]  
 RN [853]  
 RN [854]  
 RN [855]  
 RN [856]  
 RN [857]  
 RN [858]  
 RN [859]  
 RN [860]  
 RN [861]  
 RN [862]  
 RN [863]  
 RN [864]  
 RN [865]  
 RN [866]  
 RN [867]  
 RN [868]  
 RN [869]  
 RN [870]  
 RN [871]  
 RN [872]  
 RN [873]  
 RN [874]  
 RN [875]  
 RN [876]  
 RN [877]  
 RN [878]  
 RN [879]  
 RN [880]  
 RN [881]  
 RN [882]  
 RN [883]  
 RN [884]  
 RN [885]  
 RN [886]  
 RN [887]  
 RN [888]  
 RN [889]  
 RN [890]  
 RN [891]  
 RN [892]  
 RN [893]  
 RN [894]  
 RN [895]  
 RN [896]  
 RN [897]  
 RN [898]  
 RN [899]  
 RN [900]  
 RN [901]  
 RN [902]  
 RN [903]  
 RN [904]  
 RN [905]  
 RN [906]  
 RN [907]  
 RN [908]  
 RN [909]  
 RN [910]  
 RN [911]  
 RN [912]  
 RN [913]  
 RN [914]  
 RN [915]  
 RN [916]  
 RN [917]  
 RN [918]  
 RN [919]  
 RN [920]  
 RN [921]  
 RN [922]  
 RN [923]  
 RN [924]  
 RN [925]  
 RN [926]  
 RN [927]  
 RN [928]  
 RN [929]  
 RN [930]  
 RN [931]  
 RN [932]  
 RN [933]  
 RN [934]  
 RN [935]  
 RN [936]  
 RN [937]  
 RN [938]  
 RN [939]  
 RN [940]  
 RN [941]  
 RN [942]  
 RN [943]  
 RN [944]  
 RN [945]  
 RN [946]  
 RN [947]  
 RN [948]  
 RN [949]  
 RN [950]  
 RN [951]  
 RN [952]  
 RN [953]  
 RN [954]  
 RN [955]  
 RN [956]  
 RN [957]  
 RN [958]  
 RN [959]  
 RN [960]  
 RN [961]  
 RN [962]  
 RN [963]  
 RN [964]  
 RN [965]  
 RN [966]  
 RN [967]  
 RN [968]  
 RN [969]  
 RN [970]  
 RN [971]  
 RN [972]  
 RN [973]  
 RN [974]  
 RN [975]  
 RN [976]  
 RN [977]  
 RN [978]  
 RN [979]  
 RN [980]  
 RN [981]  
 RN [982]  
 RN [983]  
 RN [984]  
 RN [985]  
 RN [986]  
 RN [987]  
 RN [988]  
 RN [989]  
 RN [990]  
 RN [991]  
 RN [992]  
 RN [993]  
 RN [994]  
 RN [995]  
 RN [996]  
 RN [997]  
 RN [998]  
 RN [999]  
 RN [1000]

DR InterPro; IPR000001; Kringle.  
 DR InterPro; IPR001254; Peptidase S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR Pfam; PF00008; EGF\_1.  
 DR Pfam; PF00039; fn1\_1.  
 DR Pfam; PF00051; Kringle\_2.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00018; KRINGLE.  
 DR ProDom; PD000395; Kringle\_2.  
 DR SMART; SM00181; EGF\_1.  
 DR SMART; SM00058; FN1\_1.  
 DR SMART; SM00130; KR\_2.  
 DR SMART; SM00020; TRYD\_SPC\_1.  
 DR PROSITE; PS00022; EGF\_1.  
 DR PROSITE; PS01186; EGF\_2.  
 DR PROSITE; PS00026; EGF\_3.  
 DR PROSITE; PS01253; FIBRONECTIN\_1.  
 DR PROSITE; PS00021; KRINGLE\_1.  
 DR PROSITE; PS00070; KRINGLE\_2.  
 DR PROSITE; PS02040; TRYPSIN\_DOM.  
 DR PROSITE; PS00134; TRYPSIN\_HIS.  
 DR PROSITE; PS00135; TRYPSIN\_SER.  
 DR Plasminogen activation; Hydrolyase; Serine protease; Glycoprotein;  
 KW Plasma; Kringle; EGF-like domain; Repeat; Signal.  
 FT SIGNAL 1 17 PROBABLE.  
 FT PROPEP 18 29  
 FT CHAIN 30 559  
 FT CHAIN 30 308  
 FT CHAIN 309 559  
 FT CHAIN 36 78  
 FT DOMAIN 79 117  
 FT DOMAIN 124 205  
 FT DOMAIN 213 294  
 FT DOMAIN 309 559  
 FT ACT\_SITE 355 355  
 FT ACT\_SITE 404 404  
 FT ACT\_SITE 510 510  
 FT DISULFID 38 68  
 FT DISULFID 66 75  
 FT DISULFID 83 94  
 FT DISULFID 88 105  
 FT DISULFID 107 116  
 FT DISULFID 124 205  
 FT DISULFID 145 187  
 FT DISULFID 176 200  
 FT DISULFID 213 294  
 FT DISULFID 234 276  
 FT DISULFID 265 289  
 FT DISULFID 297 428  
 FT DISULFID 340 356  
 FT DISULFID 348 417  
 FT DISULFID 442 516  
 FT DISULFID 474 490  
 FT DISULFID 506 534  
 FT CARBOHYD 149 149  
 FT CARBOHYD 481 481  
 FT CONFLICT 380 380  
 SQ SEQUENCE 559 AA; 62903 MW; 7DBD3809C1D1C921 CRC64;  
 Query Match 24.08; Score 757.5; DB 1; Length 559;  
 Best Local Similarity 32.5%; Pred. No. 4e-46;  
 Matches 202; Conservative 86; Mismatches 192; Indels 141; Gaps 29;  
 QY 8 LHVLLMALVYKTCAG--FSLMSLLSLDPDWTTPDQDY-----SYEDYNQEEHTSS 57  
 DB 6 LCVLL-----CGVAFTL-----PDQGIHRRFRGARSYRATCRDQOT 45  
 QY 58 TLTHAENPDW---YYEDQADQCPNP-----CEHGDC--LVHGSTF 95  
 DB 46 --TYQHQSWLRPMLRGNRVYCRKNSGLAQCHSVPRSCSPRCFNGTCQALFFSDF 103

QY 96 TSCCLAPFSGNCKQ-KVQNTCKNDNPGCGQCLITQSPYYRCVCKHYTPGSC----- 147  
Db 104 VQCPDGFVGRKCDITRATCFE--GQ3--IT-----YRGTWSTAENCAECINWNSSA 152  
QY 148 --SOVVPVCRPNPCGATCSHKRSKFTACPDQ-----FX-GKP-----C 187  
Db 153 LQCKPYSARRPNAIKLG--LGNHN-----YCNEDPDVKNPCYVFKAGKVTTFPCSTPAC 205  
QY 188 EIG-SDDCVVDGYSYRGKONETVNOHACLWNHLLLOENYNNFMEDAEHTGIGEHFNC 246  
Db 206 PKGPTEDCVGKGVYRGTHSTFTSKASCLPWNMILIGKTYTAWRANSOALGLGRHNYC 265  
QY 247 RNPDADEKWCFCIKVNDKVKVEYCDVSAQAQVAYPEESPTPEKTLPGFDSGCGTEI 306  
Db 266 RNPEDGAKPCHV-NKDRKLWYCDMSPCS-----TCGLRQY 302  
QY 307 ARKIKRIYGGFKSTAGKHPWQASLOSLDLTISNPGHFGGGLIHPWCWILRAHC--T 364  
Db 303 KQOPF-RIKGSLFTDITSHFWQAAI--FVQKRSFGERFLCGGLVLSICWVLSAACHFE 359  
QY 365 DIKTRHLKVLGQDLKKEEFHQSFVEKIPKYSHYNERDEIPHNDIALKLPVDPGHC 424  
Db 360 RPPPHLKVGLGRTYRVFVEEEOFEIEKIVHKEFD--DDTYNDIALQLRSDSQC 417  
QY 425 ALESKYVTVCLPDG--SPPSGSECHISGWGVTETGK--SRQLLDKVKLIANTLCNSR 480  
Db 418 AOESSVSGTACLPDVPQVLPDTECELSGYGKHEASPPFSDRLKEARVLPSSRCTSQ 477  
QY 481 QLYDHMDIDSMICAGNLQKFG---QDTCCQSDSGGLPCEKDGTVYVYGVISWGLECGKR 536  
Db 478 HLFNKITITNMLCAGDTRTGQNDVHDACQSDSGGLVCMIDKMTLLGIISWGLGCGQK 537  
QY 537 --PGVTVQTKFLNWKATIK 555  
Db 538 DVEGIYTKVTNVLNWIQDNMK 558

RESULT 6  
TPA MOUSE  
ID TPA\_MOUSE STANDARD; PRT; 559 AA.  
AC P11214; Q91VP2;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tpa)  
DE (t-PA) (t-plasminogen activator).  
GN PIAT.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88087303; PubMed=2826484;  
RA Rickles R.J., Darrow A.L., Strickland S.;  
RT "Molecular cloning of complementary DNA to mouse tissue plasminogen  
RT activator mRNA and its expression during F9 teratocarcinoma cell  
RT differentiation";  
RL J. Biol. Chem. 263:1563-1569 (1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Mammary gland;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan E., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Scapleton M., Soares M.B., Donald M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Scherch A., Schein J.B., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
CC -!- FUNCTION: Converts the abundant, but inactive, zymogen plasminogen  
CC to plasmin by hydrolyzing a single Arg-Val bond in plasminogen. By  
CC controlling plasmin-mediated proteolysis, it plays an important  
CC role in tissue remodeling and degradation, in cell migration and  
CC many other physiological events.  
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC plasminogen to form plasmin.  
CC -!- SUBUNIT: Heterodimer of chain A and chain B held by a disulfide  
CC bond.  
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular.  
CC -!- PTM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER  
CC PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER  
CC ARG-308 CATALYZED BY PLASMIN. TISSUE KALLIKREIN OR FACTOR XA.  
CC -!- MISCELLANEOUS: Binds to the kringle structure of the fibrin A  
CC chain. Binding to fibrin enhances its catalytic activity.  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -!- SIMILARITY: Contains 1 fibronectin type I domain.  
CC -!- SIMILARITY: Contains 2 kringle domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL, J03520; AAA04070.1; --  
CC EMBL, BC011256; AAH11256.1; --  
CC EIR, A29841; A29841.  
CC HSP, P00750; IASH.  
CC MEROPS, S01.232; --  
CC MGD, MGI:97610; Plat.  
CC InterPro; IPR009003; Cys Ser trypsin.  
CC InterPro; IPR006209; EGF-like.  
CC InterPro; IPR000083; Fibnctnl.  
CC InterPro; IPR006210; IEGF.  
CC InterPro; IPR000001; Kringle.  
CC InterPro; IPR001254; Peptidase\_S1.  
CC InterPro; IPR001314; Peptidase\_S1A.  
CC Pfam; PF00008; EGF; 1.  
CC Pfam; PF00039; fn1; 1.  
CC Pfam; PF00051; kringle; 2.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PR00722; CHYMOTRYPSIN.  
CC PRINTS; PR00018; KRINGLE.  
CC ProDom; PD000395; Kringle; 2.  
CC SMART; SM00181; EGF; 1.  
CC SMART; SM00058; FN1; 1.  
CC SMART; SM00130; KR; 2.  
CC SMART; SM00020; Tryp\_SPC; 1.  
CC PROSITE; PS00022; EGF\_1; 1.  
CC PROSITE; PS01186; EGF\_2; 1.  
CC PROSITE; PS50026; EGF\_3; 1.  
CC PROSITE; PS01253; FIBRONECTIN\_1; 1.  
CC PROSITE; PS00021; KRINGLE\_1; 2.  
CC PROSITE; PS00070; KRINGLE\_2; 2.  
CC PROSITE; PS0240; TRYPsin\_DOM; 1.  
CC PROSITE; PS00134; TRYPsin\_HIS; 1.  
CC PROSITE; PS00135; TRYPsin\_SER; 1.  
KW Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;

DB 452 ASSFFSDRLKEAHRVLYPSRCSQHLFNKVTNNVLCAGDRSCGNQLDHDACCGDSG 51

OY 511 GELTCEKDGTYYYGVISWGLEGCKR--PGVYTQTKFLNWKATIK 555  
||||| : ||||| : ||||| : ||||| : ||||| :  
DB 512 GPLVMINKMTLTGIISWGLGCQKDPGVGYTKVTNYLDWIHDMX 558

RESULT 7

FA12\_HUMAN STANDARD; PRT; 615 AA.

ID ID FA12\_HUMAN STANDARD; PRT; 615 AA.  
AC AC P00748; P78339;  
DT DT 21-JUL-1986 (Rel. 01, Created)  
DT DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE DE Coagulation factor XII precursor (EC 3.4.21.38) (Hageman factor)  
DE DE (HAIF).  
GN GN Fl2.  
OS OS Homo sapiens (Human).  
OC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX OX NCBI\_TaxID=9606;  
RN RN [1]  
SEQUENCE FROM N.A.  
RP MEDLINE=88007593; PubMed=2888762;  
RX Cool D.B., McGillivray R.T.A.;  
RA Rajkumar N., Toth E.J., Yi Q., Nickerson D.A.;  
RN Submitted (AUG-2002) to the EMBL/GenBank/DBSJ databases.  
[3]  
SEQUENCE OF 4-615 FROM N.A.  
RP MEDLINE=86176794; PubMed=3754331;  
RX Tripodi M., Citarella F., Guida S., Galeffi P., Fantoni A.,  
RA Cortese R.;  
RN "cDNA sequence coding for human coagulation factor XII (Hageman).";  
NL Nucleic Acids Res. 14:3146-3146(1986).  
[4]  
SEQUENCE OF 14-615 FROM N.A.  
RP MEDLINE=86033830; PubMed=3877053;  
RX Cool D.E., Edgell C.-J.S., Louis G.V., Zoller M.J., Brayer G.D.,  
RA McGillivray R.T.A.;  
RN "Characterization of human blood coagulation factor XII cDNA.  
Prediction of the primary structure of factor XII and the tertiary  
structure of beta-factor XIIa.";  
RL J. Biol. Chem. 260:13666-13676(1985).  
[5]  
SEQUENCE OF 146-615 FROM N.A.  
RP MEDLINE=86216049; PubMed=3011063;  
RX Que B.G., Davie E.W.;  
RN "Characterization of a cDNA coding for human factor XII (Hageman  
factor)."  
RL Biochemistry 25:1525-1528(1986).  
[6]  
SEQUENCE OF 20-379.  
RP MEDLINE=85182674; PubMed=388654;  
RX McMullen B.A., Fujikawa K.;  
RN "Amino acid sequence of the heavy chain of human alpha-factor XIIa  
activated Hageman factor)."  
RL J. Biol. Chem. 260:5328-5341(1985).  
[7]  
SEQUENCE OF 354-362 AND 373-615.  
RP MEDLINE=83291041; PubMed=6604055;  
RX Fujikawa K., McMullen B.A.;  
RN "Amino acid sequence of human beta-factor XIIa.";  
RL J. Biol. Chem. 258:10924-10933(1983).  
[8]  
SEQUENCE OF 561-615 FROM N.A.  
RP TISSUE=Blood;

DR	HSPF; P00763; IDPO.
DR	MEROPS; S01_211; "
DR	Gensw; HGNC:3530; F12.
DR	MIM; 234000;
DR	GO; GO:0003805; F:blood coagulation factor XI activity; TAS.
DR	GO; GO:0003806; F:blood coagulation factor XII activity; TAS.
DR	GO; GO:0008236; F:serine-type peptidase activity; TAS.
DR	GO; GO:0007936; P:blood coagulation; TAS.
DR	InterPro; IPRO09003; Cys Ser trypsin.
DR	InterPro; IPRO06209; EGF-like.
DR	InterPro; IPRC00083; Fibrinctn1.
DR	InterPro; IPR000562; FN Type_II.
DR	InterPro; IPR000001; kringle.
DR	InterPro; IPR001254; peptidase_S1.
DR	InterPro; IPR001314; peptidase_S1A.
DR	Pfam; PF00008; EGF; 2.
DR	Pfam; PF00039; fn1; 1.
DR	Pfam; PF00040; fn2; 1.
DR	Pfam; PF00051; kringle; 1.
DR	Pfam; PF00089; trypsin; 1.
DR	PRINTS; PRO07722; CHYMOTRYPSIN.
DR	PRINTS; PRO07123; FNTYPEII.
DR	PRINTS; PRO00118; KRINGLE.
DR	PRINTS; PRO00995; FN_type_II; 1.
DR	ProDom; PD000395; Kringle; 1.
DR	SMART; SM0018; EGF; 2.
DR	SMART; SM00058; FN1; 1.
DR	SMART; SM00059; FN2; 1.
DR	SMART; SM00130; KR; 1.
DR	SMART; SM00020; Tryp_Spc; 1.
DR	PROSITE; PS00022; EGF_1; 2.
DR	PROSITE; PS01186; EGF_2; 1.
DR	PROSITE; PS00026; EGF_3; 2.
DR	PROSITE; PS01253; FIBRONECTIN_1; 1.
DR	PROSITE; PS00023; FIBRONECTIN_2; 1.
DR	PROSITE; PS00021; KRINGLE_1; 1.
DR	PROSITE; PS00070; KRINGLE_2; 1.
DR	PROSITE; PS00240; TRYPSIN_DOM; 1.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS00135; TRYPSIN_SER; 1.
KW	Glycoprotein; Blood coagulation; Plasma; Kringle; Serine protease
KW	Hydrolase; Fibrinolysis; Signal; EGF-like domain; Repeat; Zymogen
KW	Polymorphism; Disease mutation.
FT	SIGNAL 1 19
FT	CHAIN 20 372
FT	CHAIN 373 615 ALPHA-FACTOR X1IA HEAVY CHAIN.
FT	CHAIN 354 362 ALPHA-FACTOR X1IA LIGHT CHAIN.
FT	CHAIN 373 615 BETA-FACTOR X1IA PART 1.
FT	CHAIN 47 88 BETA-FACTOR X1IA PART 2.
FT	DOMAIN 94 131 FIBRONECTIN TYPE-II.
FT	DOMAIN 133 173 EGF-LIKE 1.
FT	DOMAIN 174 210 EGF-LIKE 2.
FT	DOMAIN 217 295 KRINGLE.
FT	DOMAIN 296 349 SERINE RICH.
FT	DOMAIN 373 615 SERINE PROTEASE.
FT	CARBOHYD 109 109 O-LINKED (FUC).
FT	CARBOHYD 249 249 N-LINKED (GLNAC. ).
FT	CARBOHYD 299 299 O-LINKED (POTENTIAL).
FT	CARBOHYD 305 305 O-LINKED (POTENTIAL).
FT	CARBOHYD 308 308 O-LINKED (POTENTIAL).
FT	CARBOHYD 328 328 O-LINKED (POTENTIAL).
FT	CARBOHYD 329 329 O-LINKED (POTENTIAL).

Query March 23 5%; Score 741; DB 1; Length 615;  
 Best local similarity 31.3%; Pred No 64e-45;  
 Matches 206; Conservative 79; Mismatches 210; Indels 164; Gaps 1

QY	8 LHVLLLMALVGKTAGCFSLMSLSLD-PKW-TPDQDYSDYD-----ADPC-QPNCEHGSG
DB	1 MRALLL-----GFLIVSLSTLTTPPWAFKEHKYAEHTVLTLVTGTGPCH
QY	51 QENSTSLTHAENP---DWYYT-----EDQ-----ADPC-QPNCEHGSG



```
Db 53 YHRLQYHKTHKGRPGPQPCWATTENFDQDQWGYCLEPKVKVCHSKHSPQCKGTCVN 112
Qy 91 HGSFTCSCLAPFSKCKQKQVNTCKDNPCGEGCL-----ITQSPYVRCVCKH 140
Db 113 MFSGHCLCPQHLTGNHCK-----EKCFEPQLRFFHKNWYRTEQAARCCQKG 165
Qy 141 PYTGSCSQV--PVRPNPCQNGATC---SRHKRSKFTCAPOQKGFCEIGSD--DCY 195
Db 166 P--DAHCQLASQACRTNPLGGRGLEVEGHR-----LCHCPVGYTGFCDVDTKASY 218
Qy 196 VGDGYSYRGKMTVQNHACLYWNSHLLLOENYMFEMEDAETHGIGEHNFENRNDAEKP 255
Db 219 DGRGLSYRGLARTTISGAPCQWASEATYR---NVTAEQARNWGLGCHAFCRNPNDIRP 275
Qy 256 MCFIKVTNDKXWEYCDVSACSAQOVAYPEESPT-----EPSTKL 295
Db 276 MCFV-LNRDRLSWECDLAQCCTPQAAP---PTVPSRLHVLMPAQAPKPKQPIRT 331
Qy 296 P-----GFDSCG-KTEIAERKIRYGGFKSTAGKHQWASLQ 332
Db 332 PPSQTPGALPAKREQPFSLTENGFLSCGRLKSLSSMTRVVGLVALRGHAHYIAYL 391
Qy 333 SSLPLTISPMOQH-FCGGALHPCWLVTAHCTDK--TRHLKVVLGQDCLKK-----EE 384
Db 392 -----WGSFCAGSLIAPCWLVTRAHCLQDRPAPEDLTVVLGQERRNHSCEPQT 441
Qy 385 PHEQSRVEKIPKYSHYNERDIPNDIALKL-KPVDGHCALSKYKVTCLPDGSP-P 442
Db 442 LAVRSYRLHEAFSPVSY-----QHDLLALLRLQEDADGSCALLSPYVQVCLPSGAARP 494
Qy 443 SGSE-CHISGMGVTGKSRQ-----LIDAKVLIANTLCSNRQLYDHWIDDSMTCAGNL 497
Db 495 SETTLCQVAGWG--HQFEAEYASFLQEAQVFLSLERCSADPVHGSILLPGMLCAGFL 552
Qy 498 QKPGQDTCQDGGGLTCEKDTG---YVYVIGVSMGLECGKE--EGVYTVTKFLNWK 551
Db 553 E-GTDCACQDGGGLVCDQEAERRLTQGIISWGGCGDRKPKGVYTVDAVYLAWIR 610

RESULT 8
TPA_BOVIN
ID TPA_BOVIN STANDARD; PRT; 566 AA.
AC Q28198;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Tissue-type plasminogen activator precursor (EC 3.4.21.68) (tPA)
GN PLAT.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Ravn P., Berglund L., Petersen T.E.;
RT "Cloning and characterization of the bovine plasminogen activators uPA
Rt and tPA.";
RL Int. Dairy J. 5:605-617(1995).
CC -|- FUNCTION: Converts the abundant, but inactive, zymogen plasminogen
CC to plasmin by hydrolyzing a single Arg-Val bond in plasminogen. By
CC controlling plasmin-mediated proteolysis, it plays an important
CC role in tissue remodeling and degradation, in cell migration and
CC many other physiopathological events.
CC -|- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in
CC plasminogen to form plasmin.
CC -|- SUBUNIT: Heterodimer of chain A and chain B held by a disulfide
CC bond.
CC -|- SUBCELLULAR LOCATION: secreted; extracellular.
CC -|- PTM: THE SINGLE CHAIN, ALMOST FULLY ACTIVE ENZYME, CAN BE FURTHER
```

```
CC PROCESSED INTO A TWO-CHAIN FULLY ACTIVE FORM BY A CLEAVAGE AFTER
CC ARG-314 CATALYZED BY PLASMIN, TISSUE KALLIKREIN OR FACTOR XA.
CC -|- MISCELLANEOUS: Binds to the kringle structure of the fibrin A
CC chain. Binding to fibrin enhances its catalytic activity.
CC -|- SIMILARITY: Belongs to peptidase family S1.
CC -|- SIMILARITY: Contains 1 EGF-like domain.
CC -|- SIMILARITY: Contains 1 fibronectin type I domain.
CC -|- SIMILARITY: Contains 2 kringle domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X85800; CAA59795.1; -
CC HSP; P00750; IRTF.
CC MEROPS; S01.232; -
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR06209; EGF-like.
CC InterPro; IPR000083; Fibinctnl.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR000001; Kringle.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00008; EGF_1.
CC Pfam; PF00039; fn1; 1.
CC Pfam; PF00051; kringle; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PROSITE; PS00018; KRINGLE.
CC PRODOM; PD000395; Kringle; 2.
CC SMART; SM00181; EGF; 1.
CC SMART; SM00058; FN1; 1.
CC SMART; SM00130; KR; 2.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS00026; EGF_3; 1.
CC PROSITE; PS01253; FIBRONECTIN_1; 1.
CC PROSITE; PS00021; KRINGLE_2; 2.
CC PROSITE; PS00070; KRINGLE_2; 2.
CC PROSITE; PS02040; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Plasminogen activation: Hydrolase; Serine protease; Glycoprotein;
CC Plasma; Kringle; EGF-like domain; Repeat; Signal.
CC SIGNAL
CC PROPEP 1 21
CC CHAIN 22 33 BY SIMILARITY.
CC CHAIN 34 566 TISSUE-TYPE PLASMINOGEN ACTIVATOR.
CC CHAIN 34 314 TISSUE-TYPE PLASMINOGEN ACTIVATOR A
CC CHAIN 315 566 TISSUE-TYPE PLASMINOGEN ACTIVATOR B
CC CHAIN 40 82 FIBRONECTIN TYPE-I.
CC DOMAIN 83 121 EGF-LIKE.
CC DOMAIN 128 209 KRINGLE 1.
CC DOMAIN 219 300 KRINGLE 2.
CC DOMAIN 315 566 SERINE PROTEASE.
CC ACT_SITE 361 361 CHARGE RELAY SYSTEM.
CC ACT_SITE 410 410 CHARGE RELAY SYSTEM.
CC ACT_SITE 517 517 CHARGE RELAY SYSTEM.
CC DISULFID 42 72 BY SIMILARITY.
CC DISULFID 70 79 BY SIMILARITY.
CC DISULFID 87 98 BY SIMILARITY.
CC DISULFID 92 109 BY SIMILARITY.
CC DISULFID 111 120 BY SIMILARITY.
CC DISULFID 128 209 BY SIMILARITY.
CC DISULFID 149 191 BY SIMILARITY.
CC DISULFID 180 204 BY SIMILARITY.
CC DISULFID 219 300 BY SIMILARITY.
```

FT DISULFID 240 282 BY SIMILARITY.  
FT DISULFID 271 295 BY SIMILARITY.  
FT DISULFID 303 434 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 346 362 BY SIMILARITY.  
FT DISULFID 354 423 BY SIMILARITY.  
FT DISULFID 448 523 BY SIMILARITY.  
FT DISULFID 480 496 BY SIMILARITY.  
FT DISULFID 513 541 BY SIMILARITY.  
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 487 487 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 566 AA; 63701 MW; 2EB6BEB4E32276C3 CRC64;  
  
Query Match 23.3%; Score 734.5; DB 1; Length 566;  
Best Local Similarity 34.9%; Pred. NO. 1.7e-44;  
Matches 182; Conservative 66; Mismatches 179; Indels 92; Gaps 22;  
  
QY 77 CQNPCEHGDC--LVHGTFCTCCLAPFSGNKKQ-KVQNTC-KNPFQ-RCQCLITGSP 131  
Db 87 CSEPCFNGTCTCQALYSDFVQCPEGFMGLCEIDATATCYQGVAYRGTWSTAESG 146  
QY 132 PYTRC-----VCKHPYTGSCSQVVCVREPNPCGATCSRHKRSKFTCACPDQ--- 181  
Db 147 A--ECANWNSGGLAMPYSGR-----RPNALSLG--LGNHN-----YCRNPDQDSK 188  
QY 182 -----PK-----GKFCBEGS-----DDCVGDDGYSYRGKMNRTVNOHACLYWNSHLL 223  
Db 189 PWCYVFKAGKYISFRCSTPACAKVAEEDGDCYTGNGLAYRGTRSHTKSGASCLPWNSVFL 248  
QY 224 LQNYNMFMDAETHGHEHFNCPNPDABEKWCFIKVNDKVKWYCDVSACSADQVAY 283  
Db 249 TSKYITAWKSNAPALGLGKHCHNCPNPDGAPQWCHV-WKDRQLTWBYCDVPQCV----- 301  
QY 284 PEESPTPESTKLPQFDSGCGTETAERKIKRIVGGFKSTAGKHQWQASLQSSLPATISMPQ 343  
Db 302 -----TCGLRQY-KPQPFKGLFADITSHPWQAI--FVKNRSPCE 342  
QY 344 GHFCGGALHPWVLTAAHCTDIK--TRHLKVLVGGQDLKKEEFHQSFVEKIKPKYSHY 401  
Db 343 RFLCGGILISSCWYLSAHCQBYRPPHLLKVLGRTYRLVPGEEBQTFVEKYIHKEF 402  
QY 402 NERDEIPHNDIALKLPVGDHCALESKYVKTCLPDGS--PPSGECHISGMGVETGCK 459  
Db 403 D--DDTYNDNIALHLKSDLSLTCARESASVITCLPDASLQLPDWTCELSGVGKHES 460  
QY 460 G--SRQLLDKVKLIANTLNSQLYDHMDIDSMICAGNLQKPG-----QDTCCGDSGGP 512  
Db 461 PPFSERLKEAHLVRLYPSRSTQSLFNRTVTNNMLCAGDTRSGGDHTNLHDAQQGDSGGP 520  
QY 513 LTCEKDGTYVYGVVSGNLEGGK--PGVYTVTKPLNWK 551  
Db 521 LVCKNDHMTLVGIISWGLGCGKRDVPGVTKVNYLDWIR 561  
  
RESULT 9  
URT2 DESRO  
ID URT2 DESRO STANDARD; PRT; 477 AA.  
AC P15638;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Salivary plasminogen activator alpha 2 precursor (EC 3.4.21.68) (DSPA  
DE alpha-2) (BAT-PA) (T-plasminogen activator).  
OS Desmodus rotundus (Vampire bat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;  
OC Desmodontinae; Desmodus.  
OX NCBI\_TaxID=9430;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Salivary gland;  
RC MEDLINE=92039036; PubMed=1937019;  
EX MEDLINE=92039036; PubMed=1937019;  
RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,  
RA Alagon A., Donner P., Schleuning W.D.;

RT "The plasminogen activator family from the salivary gland of the  
RT vampire bat Desmodus rotundus: cloning and expression.";  
RL Gene 105:229-237(1991).  
[2]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Salivary gland;  
RX MEDLINE=90036867; PubMed=2509450;  
RA Gardell S.J., Duong L.T., Diehl R.E., York J.D., Hare T.R.,  
RA Register R.B., Jacobs J.W., Dixon R.A.F., Friedman P.A.;  
RA "Isolation, characterization, and cDNA cloning of a vampire bat  
RT salivary plasminogen activator.";  
RT J. Biol. Chem. 264:17947-17952(1989).  
[3]  
RN CHARACTERIZATION  
RP MEDLINE=93393059; PubMed1309059;  
RX Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,  
RA Kraetzschmar J., Haendler B., Langer G., Baldol B., Witt W.,  
RA Donner P.;  
RA "Plasminogen activators from the saliva of Desmodus rotundus (common  
RT vampire bat): unique fibrin specificity.";  
RL Ann. N.Y. Acad. Sci. 667:395-403(1992).  
CC -|- FUNCTION: Probably essential to support the feeding habits of this  
CC exclusively haematophagous animal. Probable potent thrombolytic  
CC agent.  
CC -|- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC plasminogen to form plasmin.  
CC -|- ENZYME REGULATION: Activity toward plasminogen is stimulated in  
CC the presence of fibrin I.  
CC -|- SUBUNIT: Monomer.  
CC -|- DOMAIN: The fibronectin type-I domain mediates binding to fibrin,  
CC and the kringle domain apparently mediates fibrin-induced  
CC stimulation of activity.  
CC -|- SIMILARITY: Belongs to peptidase family S1.  
CC -|- SIMILARITY: Contains 1 EGF-like domain.  
CC -|- SIMILARITY: Contains 1 fibronectin type I domain.  
CC -|- SIMILARITY: Contains 1 kringle domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M63988; AAA31593.1; -;  
DR EMBL; J05082; AAA31596.1; -;  
DR PIR; A34369; A34369.  
DR PIR; J05098; J05098.  
DR HSSP; P81119; 1A51.  
DR MEROPS; S01.232; -;  
DR InterPro; IPR009003; Cys\_ser\_trypsin.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR000083; Fibraectnl.  
DR InterPro; IPR006210; IEGF.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00008; EGF; 1.  
DR Pfam; PF00039; fnl; 1.  
DR Pfam; PF00051; kringle; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00018; KRINGLE.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00181; EGF; 1.  
DR SMART; SM00058; FNI; 1.  
DR SMART; SM00130; KR; 1.  
DR SMART; SM00020; Tryp\_Spc; 1.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS00026; EGF\_3; 1.  
DR PROSITE; PS01253; FIBRONECTIN\_1; 1.

DR PROSITE; PS00021; KRINGLE 1; 1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00334; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00335; TRYPSIN\_SER; 1.  
DR PLASMINOGEN ACTIVATION; Hydrolase; Serine protease; Glycoprotein;  
KW Kringle; EGF-like domain; Signal; Multigene family.  
FT SIGNAL 1 36 POTENTIAL.  
FT CHAIN 37 477 SALIVARY PLASMINOGEN ACTIVATOR ALPHA 2.  
FT DOMAIN 40 82 FIBRONECTIN TYPE-I.  
FT DOMAIN 83 121 EGF-LIKE.  
FT DOMAIN 128 209 KRINGLE.  
FT DOMAIN 225 477 SERINE PROTEASE.  
FT ACT\_SITE 272 272 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 321 321 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 428 428 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 42 72 BY SIMILARITY.  
FT DISULFID 70 79 BY SIMILARITY.  
FT DISULFID 87 98 BY SIMILARITY.  
FT DISULFID 92 109 BY SIMILARITY.  
FT DISULFID 111 120 BY SIMILARITY.  
FT DISULFID 128 209 BY SIMILARITY.  
FT DISULFID 149 191 BY SIMILARITY.  
FT DISULFID 180 204 BY SIMILARITY.  
FT DISULFID 214 345 BY SIMILARITY.  
FT DISULFID 257 273 BY SIMILARITY.  
FT DISULFID 265 334 BY SIMILARITY.  
FT DISULFID 359 434 BY SIMILARITY.  
FT DISULFID 391 407 BY SIMILARITY.  
FT DISULFID 424 452 BY SIMILARITY.  
FT CARBOHYD 185 185 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 398 398 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 403 403 N -> K (IN REF. 2).  
FT CONFLICT 417 417 Y -> H (IN REF. 2).  
FT CONFLICT 435 435 M -> R (IN REF. 2).  
SQ SEQUENCE 477 AA; 53719 MW; 17486555C0E5077C CRC64;

Query Match  
Best Local Similarity 35.3%; Pred. No. 2 5e-44;  
Matches 155; Conservative 74; Mismatches 170; Indels 58; Gaps 16;

QY 115 KDNPCGRQCLTQSPYYR-----CVKHPYTGSCSQVVPV--CRPNPCQNG 162  
DB 42 CRDE--KTQMIYQQESWLRPEVRKSRVHCRCR---GLAQCHTPVPKSCSELRCFNG 95  
QY 163 ATCSRHKRSKFTCAPDQFKGFCIGSD--DCYVGDGYSYRGKWRNTVQHACLYWNSH 221  
DB 96 GTCWQAASFDFVCCQPKGTGKQCEVDTHATCYKQGVYTRGTSTSGAQCIWNWSN 155  
QY 222 LLLQENYNNFMEDAETHGICEHNFERNPDADKPCFICKVNDKVKWEYCDVSACSADQV 281  
DB 156 LLTRRTYNGRRSDAITLGLGNHNYCRPNPNNSKPCVY- IKASKFTLEPCSVPCSKA-- 212  
QY 282 AYPERSPTEPKLPGDFGCKTEIAERKIKRYYGFKSTAGHPQASLQSSLPITISM 341  
DB 213 -----TCGLRKYKEPQLHST--GGLFTDITSHPWQAIFAQ--NRRSS 251  
QY 342 PQGHFCGALHPQWLTAARCTDIK--TFLKVLVLDQDLKKEEPHEQSFRVEKIFKYS 399  
DB 252 GERFLCGILLSSCVLTAARCFQERYPPQHLAVLVGRVYKPKGKEQTFFVEKIVHE 311  
QY 400 HYNERDEIPNDIALKLKPDVGHCALESKYKTVCLPDGS--FPSSGSECHISGWVET 457  
DB 312 EFD--DPTNYNDIALLLKSGSPQCAQESDSVRAICLPEANLQLPDWTCELSGYGKHKS 369  
QY 458 GKG--SRQLLDKVKLIANTLNSRQLYDHMDSDMICAGNLQK---PG-QDTCQGDG 510  
DB 370 SSFYSQLEKGVRLYPSRCSCTSKFLFKVTYNNMLCAGDTSRGIYPNVHDCQGDG 429  
QY 511 GLPTCEKDGTYVYGVISWGLECKR--PGVYQVTKFLNWKATIK 555  
DB 430 GPLVCMNDNHWTLIGIISWGVGCGEKDIPGVYTKVTNYLGNWIRDNMR 476

RESULT 10  
URTB DESRO STANDARD; PRT; 431 AA.  
ID URTB DESRO  
AC P9612L; 1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Salivary plasminogen activator beta precursor (EC 3.4.21.68) (DSPA beta).  
OS Desmodus rotundus (Vampire bat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;  
OC Desmodontinae; Desmodus.  
OX NCBI\_TaxID=9430;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Salivary Gland;  
RX MEDLINE=92039036; PubMed=1937019;  
RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P., Alagon A., Donner P., Schleuning W.D.;  
RT "The plasminogen activator family from the salivary gland of the vampire bat Desmodus rotundus: cloning and expression.";  
RL Gene 105:229-237(1991).  
RN [2]  
RP CHARACTERIZATION  
RX MEDLINE=9339059; PubMed=1309059;  
RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T., Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W., Donner P.;  
RT "Plasminogen activators from the saliva of Desmodus rotundus (common vampire bat): unique fibrin specificity.";  
RL Ann. N.Y. Acad. Sci. 667:395-403(1992).  
CC -!- FUNCTION: Probably essential to support the feeding habits of this exclusively haematophagous animal. Probable potent thrombolytic agent.  
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in plasminogen to form plasmin.  
CC -!- SUBUNIT: Monomer.  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -!- SIMILARITY: Contains 1 kringle domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M63989; AAA31594.1; -;  
CC PIR; J05059; J05059.  
CC HSSP; P98119; 1A51.  
CC MEROPS; S01\_239; -;  
CC InterPro; IPR009003; Cys\_Ser\_trypsin.  
CC InterPro; IPR006209; EGF\_like.  
CC InterPro; IPR006210; IEGF.  
CC InterPro; IPR000001; Kringle.  
CC InterPro; IPR001254; Peptidase\_S1.  
CC InterPro; IPR001314; Peptidase\_S1A.  
CC Pfam; PF00008; EGF; 1.  
CC Pfam; PF00051; kringle; 1.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PR00722; CHYMOTRYPSIN.  
CC PRINTS; PR00018; KRINGLE.  
CC ProDom; PD000395; Kringle; 1.  
CC SMART; SM00181; EGF; 1.  
CC SMART; SM00130; KE; 1.  
CC SMART; SM00020; Tryp\_Spc; 1.  
CC PROSITE; PS00022; EGF\_1; 1.  
CC PROSITE; PS01186; EGF\_2; 1.  
CC PROSITE; PS50026; EGF\_3; 1.

```
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;
KW Kringle; EGF-like domain; Signal; Multigene family.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 37 431 SALIVARY PLASMINOGEN ACTIVATOR BETA.
FT DOMAIN 37 75 EGF-LIKE.
FT DOMAIN 82 163 KRINGLE.
FT DOMAIN 179 431 SERINE PROTEASE.
FT ACT_SITE 226 226 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 275 275 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 382 382 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 41 52 BY SIMILARITY.
FT DISULFID 46 52 BY SIMILARITY.
FT DISULFID 65 74 BY SIMILARITY.
FT DISULFID 82 163 BY SIMILARITY.
FT DISULFID 103 145 BY SIMILARITY.
FT DISULFID 134 158 BY SIMILARITY.
FT DISULFID 168 299 BY SIMILARITY.
FT DISULFID 211 227 BY SIMILARITY.
FT DISULFID 219 288 BY SIMILARITY.
FT DISULFID 313 388 BY SIMILARITY.
FT DISULFID 345 361 BY SIMILARITY.
FT DISULFID 378 406 BY SIMILARITY.
FT CARBOHYD 139 139 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 431 AA; 48221 MW; 699B5E675B162CBF CRC64;

Query Match
Best Local Similarity 36.0%; Score 720.5; DB 1; Length 431;
Matches 157; Conservative 73; Mismatches 159; Indels 47; Gaps 13;

QY 134 YRCVCKHPYTGFSQVVPVCRPNPCQNGATCSRHKRSKFTACPDQFGKFCETGSD- 152
D 28 YRQLARSPRAYGCGSEL-----RCFNGGTCWAASFSDFVCCPKGYTGKQCEVDTHA 80
QY 193 DCYVGDGYSGYGRKMRNTVQACLYWNSHLLQNYNMFVEDATHGIGHNFCRPDAD 252
D 81 TCYKDOGVYTRGTWSTSSGACQCNWNSNLTRETYNGRRSDAITLGLGNHNYCRPN 140
QY 253 EKPWCFIKVNDKVKWEYCDVSAGSAQDVAYPEESPTPEKLPFGFSCGKTEIAERKIK 312
D 141 SKPWCVY-IKASKFILECFSPVCSKA-----TCGLRKYKEQLH 179
QY 313 RIYGGFKTAGKHPQWASLOSPLTITSMPOGHFCGGALHPWCWILTAHCTDIK--TRH 370
D 180 ST-GGLFTDITSHPWQAIFAQ--NRRSSGERFLCGGILSSCWLTAAHCFQERYPPQH 236
QY 371 LKAVLGDQDLKKEEFHEQSERVEKIFKYSYNERDEIPHNDIALLLKLPVDGHCALESKY 430
D 237 LRVLGRYVRVPGKEEQTFEVEKCIIEEPD--DDTYNDIALLLQKSGSPQCAQESDS 294
QY 431 VKTVCPLDGS--FPSGSECHISGWGVTGTGK--SRQLLDKAVKLIANTLNSQLYDHM 486
D 295 VRAICLPEANLQLPDWTCELSGYGKHGSSSPFYSEQLKEGHRVLPSSRCTSKFLFNKT 354
QY 487 IDDSMTCAGNLQK----PG-QPTCCGDSGGPLTCEKDGTYVYGVVSGLECGKR--PGV 539
D 355 VTNNMLCAGTSGEYIPNVHDAQCQDSGGPLVCVNDNHNILLGIISWVGCGEKDIEGV 414
QY 540 YTVQTKFLNWKIKATIK 555
D 415 YTKVTNVLGWIRDNR 430

RESULT 11
ID FA12_BOVIN
AC P98140;
DT 01-FEB-1996 (Rel. 33, Created)
```

DR ProDom; PD000395; Kringle; 1.  
 DR SMART; SM00181; EGF; 2.  
 DR SMART; SM00059; FN1; 1.  
 DR SMART; SM00059; FN2; 1.  
 DR SMART; SM00130; KR; 1.  
 DR SMART; SM00023; Tryp\_Spc; 1.  
 DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; FALSE\_NEG.  
 DR PROSITE; PS00026; EGF\_3; 2.  
 DR PROSITE; PS01253; FIBRONECTIN\_1; 1.  
 DR PROSITE; PS00023; FIBRONECTIN\_2; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 1.  
 DR PROSITE; PS00070; KRINGLE\_2; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Glycoprotein; Blood coagulation; Plasma; Kringle; Serine protease;  
 KW Hydrolyase; Fibrinolysis; EGF-like domain; Repeat; Zymogen; Signal.  
 FT NON\_TER 1  
 FT SIGNAL 1  
 FT CHAIN 10 349  
 FT CHAIN 350 593  
 FT DOMAIN 37 78  
 FT DOMAIN 84 121  
 FT DOMAIN 123 163  
 FT DOMAIN 164 200  
 FT DOMAIN 207 287  
 FT DOMAIN 297 333  
 FT DOMAIN 350 593  
 FT ACT\_SITE 389 399  
 FT ACT\_SITE 438 438  
 FT ACT\_SITE 541 541  
 FT DISULFID 88 100  
 FT DISULFID 94 109  
 FT DISULFID 111 120  
 FT DISULFID 125 153  
 FT DISULFID 151 160  
 FT DISULFID 168 179  
 FT DISULFID 173 188  
 FT DISULFID 190 199  
 FT DISULFID 207 287  
 FT DISULFID 230 263  
 FT DISULFID 258 282  
 FT DISULFID 336 463  
 FT DISULFID 374 390  
 FT DISULFID 382 452  
 FT DISULFID 413 416  
 FT DISULFID 479 547  
 FT DISULFID 510 526  
 FT DISULFID 537 568  
 FT CARBOHYD 99  
 FT CARBOHYD 241  
 FT CARBOHYD 263 263  
 FT CARBOHYD 410 410  
 FT SEQUENCE 593 AA; 65148 MW; 721592BA792BD61F CRC64;

Query Match 22.4%; Score 706.5; DB 1; Length 593;  
 Best local similarity 33.0%; Pred. No. 1.7e-42;  
 Matches 191; Conservative 69; Mismatches 183; Indels 135; Gaps 31;  
 QY 56 SSTLTHAENPDWYV---TEQADPC-QPNCEHGGDCLVHGSFTTCSCLAPFSGNKCKQV 111  
 DB 64 ATTFNEKQORWAYCLEPKVKDCHSKNDCQGGTCVMPDGPICADHFTGKHCQK- 122  
 QY 112 QNTCKDNPGRGQCLITQSPYYR-----CVCKHP--YTGPCSQVY 151  
 DB 123 -----EKCF---EPQFFRPHENBIHRLBPAGVVKCQKGPNAQCKPLASQ-- 166  
 QY 152 PVCRPNPCQNGATCSR---HKRSKFTCACPDQFKGKFCBIG-SDCY--VGDGYGRGK 205  
 DB 167 -VKTNEFLNGDSCLEQEGH-----LCRCAPFAGRLCDVDLKASCYDDRDLGLYRGM 220  
 QY 206 MNRVTNQHACLYWNHLLQENYMFMDAETHGIGHNFCRNPDADEKPCWCFIKVTNDK 265

DB 221 AGTTLGAPCQWASEATY---WNVTAQVLNWLGDHAFCEPNDNTEPWCFI-WKGD 276  
 QY 266 VRWEYCDVSAC--SAQDVAYPEESP-----TESTKLPGRDS--CGKTEIAE----- 308  
 DB 277 LSNWYCRLLAPCAAAGHEHFLPSPSALQKPESTTTQTPLPSTSGWCSPTPLASGPGGC 336  
 QY 309 -----RK-----IKRIYGGFKSTAGKHEWQASLQSSPLTITSMFQHPGCGGALHPCWVLT 360  
 DB 337 GQRLRWLSSLNRVVGGLVALPGAHYIALYN-----DQHFCASSLAPCWVLT 387  
 QY 361 AHCTDIK--TRHLKVVGLGQDLKKEEFHQS-----FRVEKIFK---YSHYNER 404  
 DB 388 AHCLQNRPAKELTVVLG-----QDRHQSCQECQCTLAVRDYRLHEAFSPITYQH---- 437  
 QY 405 DEIPHNDILLKL-KPVDGHCALESKYVTVCLPD--GSFPGSE-----CHISGWG-VTE 458  
 DB 438 -----DLALVRLQESADGCCAPSPFVQVCLPSTPAARPAESEAACEVAGHGFEGG 491  
 QY 459 KGSRLQLDAKVKLIANTLNSRLYDHMDMSICAGNLQKPGQDTCQDSSGGPLTCE-- 516  
 DB 492 EYSSFLOEAQVPLIDPQRCAPDVHGAFTQGLCAGFLE-GGTDACQDSSGGPLTCEDE 550  
 QY 517 -KGTYYVYGVISWGLECGKR--PGVYTVQTKELNWK 551  
 DB 551 TPERQLILRGIVSWGSGCGNRLKPGVYTDVANYLAWIR 588  
 RESULT 12  
 URTL\_DESRO STANDARD; PRT; 477 AA.  
 AC P98119;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Salivary plasminogen activator alpha 1 precursor (EC 3.4.21.68) (DSPA  
 alpha-1).  
 OS Desmodus rotundus (Vampire bat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Phyllostomidae;  
 OC Desmodontinae; Desmodus.  
 OX NCBI\_TaxID=9430;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Salivary gland;  
 RX MEDLINE=92039036; PubMed=1937019;  
 RA Kraetzschmar J., Haendler B., Langer G., Boidol W., Bringmann P.,  
 Alagon A., Donner P., Schleuning W.D.;  
 RT "The plasminogen activator family from the salivary gland of the  
 vampire bat Desmodus rotundus: cloning and expression.";  
 RL Gene 105:229-237(1991).  
 RN [2]  
 RP CHARACTERIZATION.  
 RC MEDLINE=93393059; PubMed=1309059;  
 RA Schleuning W.-D., Alagon A., Boidol W., Bringmann P., Petri T.,  
 Kraetzschmar J., Haendler B., Langer G., Baldus B., Witt W.,  
 Donner P.;  
 RT "Plasminogen activators from the saliva of Desmodus rotundus (common  
 vampire bat): unique fibrin specificity.";  
 RL Ann. N.Y. Acad. Sci. 667:395-403(1992).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).  
 RC TISSUE=Salivary gland;  
 RX MEDLINE=98022741; PubMed=9354616;  
 RA Renatus M., Stubbs M.T., Huber R., Bringmann P., Donner P.,  
 Schleuning W.D., Bode W.;  
 RT "Catalytic domain structure of vampire bat plasminogen activator: a  
 molecular paradigm for proteolysis without activation cleavage.";  
 RL Biochemistry 36:13483-13493(1997).  
 CC -!- FUNCTION: Probably essential to support the feeding habits of this  
 exclusively haematophagous animal. Potent thrombolytic agent.  
 CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
 plasminogen to form plasmin.

CC -!- ENZYME REGULATION: Activity toward plasminogen is stimulated in  
CC the presence of fibrin I.  
CC -!- SUBUNIT: Monomer.  
CC -!- DOMAIN: The fibronectin type-I domain mediates binding to fibrin,  
CC and the krigle domain apparently mediates fibrin-induced  
CC stimulation of activity.  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -!- SIMILARITY: Contains 1 fibronectin type I domain.  
CC -!- SIMILARITY: Contains 1 krigle domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M63987; AAA31591.1; -  
CC EMBL; M63986; AAA31592.1; -  
CC PIR; JS0597; JS0597.  
CC PDB; 1A51; 23-MAR-99.  
CC MEROPS; S01.232; -  
CC GlycositesDB; P98119; -  
CC InterPro; IPR009003; Cys\_Ser\_trypsin.  
CC InterPro; IPR006209; EGF\_like.  
CC InterPro; IPR000083; Fibrinctn1.  
CC InterPro; IPR006210; IEGF.  
CC InterPro; IPR000001; Krigle.  
CC InterPro; IPR001254; Peptidase\_S1.  
CC InterPro; IPR001314; Peptidase\_S1A.  
CC Pfam; PF00008; EGF; 1.  
CC Pfam; PF00039; fnl; 1.  
CC Pfam; PF00051; krigle; 1.  
CC Pfam; PF00089; trypsin; 1.  
CC PRINTS; PR00722; CHYMOTRYPSIN.  
CC PRINTS; PR00018; KRINGLE.  
CC ProDom; PD000395; Krigle; 1.  
CC SMART; SM00181; EGF; 1.  
CC SMART; SM00056; FN1; 1.  
CC SMART; SM00130; KR; 1.  
CC SMART; SM00020; Tryp\_Spc; 1.  
CC PROSITE; PS00022; EGF\_1; 1.  
CC PROSITE; PS01186; EGF\_2; 1.  
CC PROSITE; PS00026; EGF\_3; 1.  
CC PROSITE; PS01253; FIBRONECTIN\_1; 1.  
CC PROSITE; PS00021; KRINGLE\_1; 1.  
CC PROSITE; PS00070; KRINGLE\_2; 1.  
CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
CC Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;  
CC Krigle; EGF-like domain; Signal; Multigene family; 3D-structure.  
KW Krigle; EGF-like domain; Signal; Multigene family; 3D-structure.  
FT SIGNAL 1 36  
FT CHAIN 37 477  
FT DOMAIN 40 82  
FT DOMAIN 83 121  
FT DOMAIN 128 209  
FT DOMAIN 225 477  
FT ACT\_SITE 272 272  
FT ACT\_SITE 321 321  
FT ACT\_SITE 428 428  
FT DISULFID 42 72  
FT DISULFID 70 79  
FT DISULFID 87 98  
FT DISULFID 92 109  
FT DISULFID 111 120  
FT DISULFID 128 209  
FT DISULFID 149 191  
FT DISULFID 180 204  
FT DISULFID 214 345  
FT DISULFID 257 273

FT DISULFID 265 334  
FT DISULFID 359 434  
FT DISULFID 391 407  
FT DISULFID 424 452  
FT CARBOHYD 153 153  
FT CARBOHYD 398 398  
FT TURN 214 215  
FT STRAND 223 224  
FT TURN 226 227  
FT STRAND 230 231  
FT HELIX 234 236  
FT TURN 238 239  
FT STRAND 240 245  
FT STRAND 254 263  
FT TURN 264 265  
FT STRAND 266 269  
FT HELIX 271 273  
FT TURN 280 282  
FT STRAND 284 287  
FT TURN 297 298  
FT STRAND 300 309  
FT TURN 311 312  
FT TURN 315 317  
FT TURN 319 320  
FT STRAND 323 328  
FT STRAND 338 338  
FT TURN 339 340  
FT STRAND 341 341  
FT STRAND 345 345  
FT TURN 349 350  
FT TURN 355 356  
FT STRAND 358 363  
FT STRAND 366 366  
FT STRAND 374 374  
FT STRAND 379 385  
FT HELIX 388 390  
FT TURN 393 398  
FT TURN 403 404  
FT STRAND 405 409  
FT TURN 425 426  
FT TURN 428 429  
FT STRAND 431 436  
FT TURN 437 438  
FT STRAND 439 448  
FT TURN 455 456  
FT STRAND 459 463  
FT HELIX 464 467  
FT HELIX 468 474  
SQ SEQUENCE 477 AA; 53616 MW; AA06FD1739C10E5E CRC64;  
Query Match 22.3%; Score 704.5; DB 1; Length 477;  
Best Local Similarity 35.9%; Pred. NO. 1.8e-42;  
Matches 161; Conservative 64; Mismatches 160; Indels 63; Gaps 15;  
QY 120 CGRGGLITQSPYVRCVKCHPYTGTGSCSQVVPV--CRPNPCQNGATCSRHKRRSKFTCA 177  
DB 72 CDRGQA-----RC-----HIVPVNSCSEPCFNGGTQWQVYFSDVVCQ 110  
QY 178 CPDQFKGKFCIGS-DDCYVGDSYRGKMRNTVNHACLYWNSHLLLOENYMFMDAE 236  
DB 111 CPAGYTGKCEVDTRATCYEGQGVYRGTWSTAESRVEICINWNSLLTRTYNGRMPDAF 170  
QY 237 THGIGEHNECRNPDADEKPMWCFIKVNDKVKWEYCDVSACSAQDVAYPEESPTPESTKLP 296  
DB 171 NLGLGNHNYCRNPNAGAPKPCVIV-IRAGKFTSESCSVPCSKA----- 212  
QY 297 GPDSCGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPLTISMPQGHFCGALIHPCW 356  
DB 213 ---TCGLRKVKPQLHST--GGLFTDITSHPWQAIFAQ--NRRSSGERFLCGGILISSCW 266  
QY 357 VLTAHHC--TDIKTRHLKVLGDQDLKKEEFHQSFQFRVEKIFKYSHYNERDEIPNDIAL 414

N-LINKED (GLCNAC. . .).  
/FTID-CAR\_000027.  
N-LINKED (GLCNAC. . .).  
/FTID-CAR\_000028.

```
Db 267 VITAARHCFQESYLPOLKVLGRVTRVVRKPEEZEQTFKVKYIVHKEFD--DDTYNDIAL 324
QY 415 LKLPVVDGHCALSKYKVTCLPDGS--FPGSECHISGWGTETGK--SRQLLDKVK 470
Db 325 LQKSDSPCAQESVSRAICLPEANLQLPDWTCELSGYGKHSSFFSEQLKEGHRV 384
QY 471 LIANTLNSRQYIDHMDSDMTCAGNLQK----PG-QDTCCGDSGGLTCCKDGTYYVG 525
Db 385 LYPSSRCAPKFLFNKVTNNMLCAGDTSRGEIYNVHDACQDGGPLVCMNDNMTLLG 444
QY 526 IVSWGLECGKR--PGVYTVQTKFLNWK 551
Db 445 IISWVGCGEKDVPYIKVNYLQWIR 472

RESULT 13
UROK RAT
ID UROK RAT STANDARD; PRT; 432 AA.
AC P29598;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Urokinase-type plasminogen activator precursor (EC 3.4.21.73) (uPA)
DE (U-plasminogen activator).
GN PLAU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Eischer 344; PubMed=1568219;
RA Henderson B.R., Tansey W.P., Phillips S.M., Ramshaw I.A.,
RA Kefford R.F.;
RT "transcriptional and posttranscriptional activation of urokinase
RT plasminogen activator gene expression in metastatic tumor cells.";
RL Cancer Res. 52:2489-2496(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Rabbani S.A.;
RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.
CC -! CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in
CC plasminogen to form plasmin.
CC -! SUBUNIT: Found in high and low molecular mass forms. Each consists
CC of two chains, A and B. The high molecular mass form contains a
CC long chain A. Cleavage occurs after residue 156 in the low
CC molecular mass form to yield a short A1 chain (By similarity).
CC -! SIMILARITY: Belongs to peptidase family S1.
CC -! SIMILARITY: Contains 1 EGF-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X63434; CAA45028.1; -
CC EMBL; X65651; CAA46601.1; -
CC FIR; S24604; S18932.
CC HSSP; P00749; 1KDU.
CC MEROPS; S01.231; -.
CC InterPro; IPR009003; Cys Ser trypsin.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR000001; Kringle.
CC InterPro; IPR006293; Pept_S1A_uPA.
CC InterPro; IPR001254; Peptidase_S1.
```

```
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PIRSF; PIRSF001144; Ufk_plasm_act; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR0018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR PROSITE; PS00026; EGF_3; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;
KW Kringle; EGF-like domain; Zymogen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 432 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.
FT CHAIN 20 177 CHAIN A (BY SIMILARITY).
FT CHAIN 156 177 SHORT A CHAIN (A1) (BY SIMILARITY).
FT CHAIN 179 432 CHAIN B (BY SIMILARITY).
FT DOMAIN 27 63 EGF-LIKE.
FT DOMAIN 70 151 KRINGLE.
FT DOMAIN 152 178 CONNECTING PEPTIDE.
FT DOMAIN 179 432 SERINE PROTEASE.
FT DISULFID 31 39 BY SIMILARITY.
FT DISULFID 33 51 BY SIMILARITY.
FT DISULFID 53 62 INTERCHAIN (BY SIMILARITY).
FT DISULFID 168 300 BY SIMILARITY.
FT DISULFID 210 226 BY SIMILARITY.
FT DISULFID 218 289 BY SIMILARITY.
FT DISULFID 314 383 BY SIMILARITY.
FT DISULFID 346 362 BY SIMILARITY.
FT DISULFID 373 401 BY SIMILARITY.
FT ACT_SITE 225 225 CHARGE RELAY SYSTEM.
FT ACT_SITE 276 276 CHARGE RELAY SYSTEM.
FT ACT_SITE 377 377 CHARGE RELAY SYSTEM.
FT CONFLICT 16 16 N -> H (IN REF. 2).
FT CONFLICT 24 24 E -> G (IN REF. 2).
FT CONFLICT 332 332 D -> N (IN REF. 2).
SQ SEQUENCE 432 AA; 47957 MW; 4EBB96C716244C8 CRC64;

Query Match 22.3%; Score 704; DB 1; Length 432;
Best Local Similarity 37.9%; Pred. No. 1.8e-42;
Matches 157; Conservative 64; Mismatches 163; Indels 30; Gaps 14;

QY 159 CONGATCSRHKRSKF-TCACPDQFKYKCFEIG-SDDCYVDGYSYRGKMNRTVNQHA 216
DB 33 CONGGVCVSYKYFSSIRRCSPKFKGEGHCEIDTSKTYHNGNQSTRGKANTDTKGRPCL 92

QY 217 YWNSHLLQENYMFMEDEATHGIEHNFRCNPDADEKPCWCFIKVTNDKYNKBYCDVSAC 276
DB 93 AWNSPAVLQTYNAHRSDALSGLGKHNYCRNPDNQRPWCYVQIGLKQF-----VQEC 146

QY 277 SAQDVAYPESPTPEPTKL--PGFDSCKGTEIAERIKRIYGFKSTAGKHQASLQSS 334
DB 147 MVQDCSLSK-----KPSSTVDQGF-QCGQKALRPR--FKIVGGEFTVENQPFALY-- 197

QY 335 LPLTISMPQGHFGGALIHPCWVLTAAHC--TDIKTRHLKVLGDDQDLKKEEFEEQSFRV 392
DB 196 LKNGGSGPPSPKCGSILSPCWASATHCFVNPQKKEEYVVVLGQSKNSYNPGEMKFEV 257

QY 393 EKIFKYSHYNERDEIPHNDIALKLKPVDDHCALESKYVKTCLPD--GSFPGSECHIS 450
DB 258 EQLIHEDFSDETLAFHNDIALKIRTSQCCQAPSTIQTICLPFRGDPAPFGDCEIT 317

QY 451 GWGVETGKG--SRQLLDKVKLIANTLNSRQLYDHMDSDMTCAGNLQKPGQDTCCG 507
DB 318 GFG-QESATDYFPKDLKMSVKIISHEQCKQPHYGYSEINYNKMLCAADPEWK-TDSCSG 375
```





OC Bovidae; Bovinae; Bos.  
ON NCBI\_TaxID=9913;  
RX MEDLINE=93216119; PubMed=8385052;  
RA Kraetzschmar J., Haendler B., Kojima S., Rifkin D.B.,  
RA Schleuning W.-D.;  
RT "Bovine urokinase-type plasminogen activator and its receptor:  
RT cloning and induction by retinoic acid.";  
RL Gene 125:177-183(1993).  
RN (2)  
RP SEQUENCE OF 12-433 FROM N.A.  
RC TISSUE=Kidney;  
RA Ravn P., Berglund L., Petersen T.B.;  
RT "Cloning and characterization of the bovine plasminogen activators uPA  
RT and tPA.";  
RL Int. Dairy J. 5:605-617(1995).  
CC -!- CATALYTIC ACTIVITY: Specific cleavage of Arg-|-Val bond in  
CC plasminogen to form plasmin.  
CC -!- INDUCTION: By retinoic acid.  
CC -!- SIMILARITY: Belongs to peptidase family S1.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -!- SIMILARITY: Contains 1 kringle domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L03546; AAA51419.1; -;  
CC EMBL; X85801; CAA59796.1; -;  
CC FIR; JN0560; JN0560.  
CC HSSP; P00749; 1LMW.  
CC MEROPS; S01.231; -;  
CC InterPro; IPR009003; Cys Ser trypsin.  
CC InterPro; IPR006209; EGF-like.  
CC InterPro; IPR000001; Kringle.  
CC InterPro; IPR008293; Pept\_S1A\_uPA.  
CC InterPro; IPR001254; Peptidase\_S1.  
CC InterPro; IPR001314; Peptidase\_S1A.  
CC Pfam; PF00051; kringle; 1.  
CC Pfam; PF00089; trypsin; 1.  
CC PIRSF; PIRSF00144; Urk\_plasm\_act; 1.  
CC PRINTS; PR00722; CHYMOTRYPsin.  
CC PRINTS; PD000395; Kringle; 1.  
CC ProDom; PD000395; Kringle; 1.  
CC SMART; SM00130; KR; 1.  
CC SMART; SM00020; Tryp\_Spc; 1.  
CC PROSITE; PS00022; EGF\_1; 1.  
CC PROSITE; PS01185; EGF\_2; FALSE\_NEG.  
CC PROSITE; PS00026; EGF\_3; 1.  
CC PROSITE; PS00021; KRINGLE\_1; 1.  
CC PROSITE; PS00070; KRINGLE\_2; 1.  
CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
CC Plasminogen activation; Hydrolase; Serine protease; Glycoprotein;  
CC Kringle; EGF-like domain; Signal; Zymogen.  
CC SIGNAL 1 20  
FT CHAIN 21 433 UROKINASE-TYPE PLASMINOGEN ACTIVATOR.  
FT CHAIN 21 179 CHAIN A (BY SIMILARITY).  
FT CHAIN 181 433 CHAIN B (BY SIMILARITY).  
FT DOMAIN 29 65 EGF-LIKE.  
FT DOMAIN 72 153 KRINGLE.  
FT DOMAIN 154 180 CONNECTING PEPTIDE.  
FT DOMAIN 181 433 SERINE PROTEASE.  
FT DISULFID 33 41 BY SIMILARITY.  
FT DISULFID 35 53 BY SIMILARITY.  
FT DISULFID 55 64 BY SIMILARITY.

FT DISULFID 170 301 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 211 227 BY SIMILARITY.  
FT DISULFID 219 290 BY SIMILARITY.  
FT DISULFID 315 384 BY SIMILARITY.  
FT DISULFID 347 363 BY SIMILARITY.  
FT DISULFID 374 402 BY SIMILARITY.  
FT ACT\_SITE 226 226 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 277 277 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 378 378 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT CONFLICT 189 189 A -> T (IN REF. 2).  
SQ SEQUENCE 433 AA; 48730 MW; 4DE1B8D4DA47027A CRC64;  
  
Query Match 21.7%; Score 683.5; DB 1; Length 433;  
Best Local Similarity 26.7%; Pred. No. 4.9e-41;  
Matches 152; Conservative 66; Mismatches 161; Indels 35; Gaps 13;  
  
QY 159 CONGATCSRHRKRSKP-TCACPDQFKGKFCIEG-SDDCVGDCGYSGKQNRVTHQACL 216  
DB 35 CLNGGKCVTKYKFSNIQRCSCKPKFQGEHCIEDTSKTCYQGNHGSYRGKANDLSGRPCL 94  
  
QY 217 YKNSHLLLOENYNMFEDAEETHGIGHNFCRNPDADEKPCFKVTNDKVKWEYCDVSAC 276  
DB 95 AWDSPTVLLKMYHAHRSDAIQLGLGKNYCRNPDRNORRWPYVQI-GLKQFVQFCVQDC 153  
  
QY 277 SAQDVAYPEESPTPESTKLPFGDSCGKTEIAERKIRIYGGFKSTAGKHPWQASLQ---- 332  
DB 154 SV-----GKSPSPREKEB--FCCGQKALRPR--FKIVGGQVTNAENQFWFAIYRRHR 203  
  
QY 333 -SSLPUTISMPQGHFCGALIHQCWLTAHC--TDIKTRHLKVLGDQDLKKEEFHQES 389  
DB 204 GGSIT-----YLCGSLISPCWVVSATHCFIDHPKKNYIYVLCQSLNSDTGEMQ 255  
  
QY 390 FRVEKIFKYSHYNERDEIPHNDIALKLKFPVDGHCALSKYKVTCLP--DGSFPPGSEC 447  
DB 256 FEVEKLILHEDYSAESLAHENDIALKIRTSRQCAQPSRSIOTICLPPEHDAHSRTRC 315  
  
QY 448 HISGNGVTETG--KGSRLLDKVKLIANTLCNSRQLYDHWDSDMTICAGNLOKPGGOTC 505  
DB 316 EITGFGKENPSDYRSDDELKMTFVSLVSEVCOQPHYGAETDKMLCAADPOWE-TDSC 374  
  
QY 506 QGDSGGPLTCEKDGTYVYVYGVSMGLECGK--RPGVYTVTKFLNMIKATIKSE 557  
DB 375 QGDSGGPLVCTIQRLTLTGIVSWGRDCAMKYKFGVYTVSKFLPWINTHTRGE 428  
  
Search completed: May 24, 2004, 09:47:26  
Job time : 13.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 09:43:35 ; Search time 39 Seconds  
(without alignments)  
4530.518 Million cell updates/sec

Title: US-09-912-559-3  
Perfect score: 3154  
Sequence: 1 MFARMSDLHVLLMALVGT.....TQVTKFLNWKATIKSESSEGF 560

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL 25:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriaph:\*
  - 17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3154	100.0	560	4 Q14520	Q14520 homo sapien
2	2363.5	74.9	517	11 Q8KOD2	Q8KOD2 mus musculus
3	831	26.3	540	13 Q800Y7	Q800Y7 meleagris g
4	829	26.3	653	11 Q8VCS4	Q8VCS4 mus musculus
5	766.5	24.3	516	4 Q8BU99	Q8BU99 homo sapien
6	765.5	24.3	562	4 Q86YK8	Q86YK8 homo sapien
7	758.5	24.0	562	6 Q8SQ23	Q8SQ23 sus scrofa
8	737	23.4	564	6 Q8MKB1	Q8MKB1 oryctolagus
9	728	23.1	615	4 Q81Z25	Q81Z25 homo sapien
10	725.5	23.0	616	6 Q97507	Q97507 sus scrofa
11	712	22.6	395	4 Q9BZW1	Q9BZW1 homo sapien
12	694	22.0	597	11 Q35727	Q35727 mus musculus
13	689	21.8	609	11 Q80YCS	Q80YCS mus musculus
14	667.5	21.2	433	6 Q8MIL0	Q8MIL0 oryctolagus
15	666.5	21.1	433	6 Q8MHY7	Q8MHY7 oryctolagus
16	641.5	20.3	810	4 Q15146	Q15146 homo sapien

17	627.5	19.9	454	6	O46506	O46506 papio hamad
18	587.5	18.6	812	11	Q9R0W3	Q9R0W3 rattus norv
19	586	18.6	806	6	O18783	O18783 macropus eu
20	586	18.6	429	13	Q8AVB0	Q8AVB0 brachydanio
21	575	18.2	334	6	O46507	O46507 papio hamad
22	506	16.0	385	5	Q25101	Q25101 herdmantia m
23	468.5	14.9	868	5	Q9YIV3	Q9YIV3 polyandroca
24	465	14.7	716	11	P70521	P70521 rattus norv
25	460	14.6	709	13	Q90ZM6	Q90ZM6 brachydanio
26	456.5	14.5	781	11	Q99JTC8	Q99JTC8 rattus norv
27	455	14.4	716	11	Q91XG8	Q91XG8 mus musculu
28	454	14.4	471	13	Q804X6	Q804X6 gallus gall
29	450.5	14.3	422	4	Q8WVC1	Q8WVC1 homo sapien
30	449	14.2	716	13	Q91691	Q91691 xenopus lae
31	446.5	14.2	461	6	Q95ND7	Q95ND7 pan troglod
32	446	14.1	581	5	Q9XZM7	Q9XZM7 strongyloce
33	444	14.1	710	13	Q91402	Q91402 xenopus lae
34	443	14.0	455	13	Q78186	Q78186 xenopus lae
35	440	14.0	717	13	P70006	P70006 xenopus lae
36	439.5	13.9	728	11	Q8C9G5	Q8C9G5 mus musculu
37	438.5	13.9	433	13	Q804X5	Q804X5 gallus gall
38	438.5	13.9	855	11	Q93J17	Q93J17 rattus norv
39	435.5	13.8	709	13	Q72TN9	Q72TN9 xenopus lae
40	433.5	13.7	425	13	Q804X7	Q804X7 gallus gall
41	433	13.7	503	13	Q8AYE4	Q8AYE4 brachydanio
42	432.5	13.7	446	11	Q8K3U6	Q8K3U6 rattus norv
43	432.5	13.7	490	11	Q7TN04	Q7TN04 mus musculu
44	432.5	13.7	624	11	Q9DAT3	Q9DAT3 mus musculu
45	432	13.7	1322	5	Q9NJ55	Q9NJ55 anopheles g

## ALIGNMENTS

RESULT 1	Q14520	PRELIMINARY;	PRT;	560 AA.
ID	Q14520	PRELIMINARY;		
AC	Q14520	O00663;		
DT	01-NOV-1996	(TrEMBLrel. 01, Created)		
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)		
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)		
DE	HGF activator like protein (Hyaluronan binding protein 2).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Kitamura N.;			
RL	Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96425001; PubMed=8827452;			
RA	Choi-Miura N.H., Tobe T., Sumiya J., Nakano Y., Sano Y., Mazda T., Tomita M.;			
RT	"Purification and characterization of a novel hyaluronan-binding protein (PHBP) from human plasma: it has three EGF, a kringle and a serine protease domain, similar to hepatocyte growth factor activator.";			
RT	J. Biochem. 119:1157-1165(1996).			
RT	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Colon, and Kidney;			
RA	Straussberg R.;			
RL	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
CC	-!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.			
DR	EMBL; D49742; BAA08576.1; -;			
DR	EMBL; S83182; AAB46909.1; -;			
DR	EMBL; BC031412; AAH31412.1; -;			
DR	PIR; JC4795; JC4795.			
DR	HSP; P00763; IDPO.			
DR	MEROPS; S01.033; -;			

```
DR Genew; HGNC:4798; HABP2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0005339; F:glycosaminoglycan binding; TAS.
DR GO; GO:0007155; P:cell adhesion; TAS.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF_3.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle.
DR ProSite; PS00022; EGF_1; 3.
DR ProSite; PS01186; EGF_2; 2.
DR ProSite; PS00021; KRINGLE 1; 1.
DR ProSite; PS00021; KRINGLE 2; 1.
DR ProSite; PS00240; TRYPSIN_DOM; 1.
DR ProSite; PS00134; TRYPSIN_HIS; 1.
DR ProSite; PS00135; TRYPSIN_SER; 1.
DR EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
SQ SEQUENCE 560 AA; 62671 MW; 5C1907230784ACD4 CRC64;

Query Match 100.0%; Score 3154; DB 4; Length 560;
Best Local Similarity 100.0%; Pred. No. 3.2e-282;
Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLVGVKTCAGFSLMSLLESLEDPDTPDQDYDYSDYEDYDQNTSSTLT 60
DB 1 MFARMSDLHVLMLVGVKTCAGFSLMSLLESLEDPDTPDQDYDYSDYEDYDQNTSSTLT 60

QY 61 HAENPDWYTTDQADPCPNCEHGGDCLVHGSTFTCSCLAPFGSGKCKQVNTCKDNPC 120
DB 61 HAENPDWYTTDQADPCPNCEHGGDCLVHGSTFTCSCLAPFGSGKCKQVNTCKDNPC 120

QY 121 GRGQCLITQSPYYRCVKHPYTGSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180
DB 121 GRGQCLITQSPYYRCVKHPYTGSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

QY 181 QFKGFCEIGSDDCVVGCGYSYRGKMTVNOHACLWNHLLQENYMFEDAEATHGI 240
DB 181 QFKGFCEIGSDDCVVGCGYSYRGKMTVNOHACLWNHLLQENYMFEDAEATHGI 240

QY 241 GEHNFERNPDADKPMWCFIKVNDKVMWEYCDVSAQSDVAYPEESTPSTKLPGDS 300
DB 241 GEHNFERNPDADKPMWCFIKVNDKVMWEYCDVSAQSDVAYPEESTPSTKLPGDS 300

QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSPLTISMPQGHFCGALIHPCWVLT 360
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSPLTISMPQGHFCGALIHPCWVLT 360

QY 361 AHCTDIKTRHLKVLGDDQLKKEEFHEQSFVEKIFKYSYHNERDEIPHNDIALLKLPV 420
DB 361 AHCTDIKTRHLKVLGDDQLKKEEFHEQSFVEKIFKYSYHNERDEIPHNDIALLKLPV 420

QY 421 DGHCALESKYVKTVCPLPDGSPGSECHTSGVGTETGKSGRQLLDKVKLIANTLNSR 480
DB 421 DGHCALESKYVKTVCPLPDGSPGSECHTSGVGTETGKSGRQLLDKVKLIANTLNSR 480

QY 481 QLYDHMDIDSMICAGNLQKPGDTCQDSDGGPLTCEKDGTYVYGVIVSGLECGKRPVY 540
DB 481 QLYDHMDIDSMICAGNLQKPGDTCQDSDGGPLTCEKDGTYVYGVIVSGLECGKRPVY 540

QY 541 TQVTKFLNWKATIKSESGF 560
DB 541 TQVTKFLNWKATIKSESGF 560

RESULT 2
Q8KOD2
```

```
ID Q8KOD2 PRELIMINARY; PRT; 517 AA.
AC Q8KOD2.
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
DE Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
EMBL; BC031775; AAH31775.1; -
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00181; EGF_3.
DR SMART; SM00130; KR_1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00021; KRINGLE 1; 1.
DR PROSITE; PS00021; KRINGLE 2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hypothetical protein; EGF-like domain; Glycoprotein; Hydrolase;
KW Kringle; Protease; Serine protease.
SQ SEQUENCE 517 AA; 57326 MW; 3855A42035A5EA59 CRC64;

Query Match 74.9%; Score 2363.5; DB 11; Length 517;
Best Local Similarity 75.1%; Pred. No. 2.6e-209;
Matches 413; Conservative 35; Mismatches 63; Indels 39; Gaps 1;

QY 10 VLLLMALVGTACGFSLSLLESLEDPDTPDQDYDYSDYEDYDQNTSSTLTAAENPDWY 69
DB 6 VLLLMALVGSVTVGLSLSFIAPDPD----- 31

QY 70 TEDQADPCPNCEHGGDCLVHGSTFTCSCLAPFGSGKCKQVNTCKDNPCRGQCLITQ 129
DB 32 -----DFQCQNPCEHGGDCLIRGDTSCSPAPFGSGRCQTAQNKCKDNPCVHGDCLITQ 86

QY 130 SPYYRCVKHPYTGSCQVVPVCRPNPCQNGATCSRHKRSKFTACPDQFKGFCEI 189
DB 87 KHPYRCACPKYPTGPDCKVLPAKPNPCQNGVCSRRHRRSRFTACPDQYKGFCEI 146

QY 190 GSDDCVVGCGYSYRGKMTVNOHACLWNHLLQENYMFEDAEATHGIGHNFCRNP 249
DB 147 GPDDCVVGCGYSYRGKMTVNOHACLWNHLLQENYMFEDAEATHGIAHNFCRNP 206

QY 250 DADKPMWCFIKVNDKVMWEYCDVSAQSDVAYPEESTPSTKLPGDSCKTKIAER 309
```

```
Db 207 DGDHKEWCFKVNSEKVKWEYCDVTVCPVDPDTPNVESSLPEVWMLPGFESCGKTEVAEH 266
Qy 310 KIKRIYGGFKSTAGKHPQASLQSSLPITISNPQHFQCGALIHPCWVLTAAHCTDIXTR 369
Db 267 AVKRIYGGFKSTAGKHPQVSLQSLPITISNPQHFQCGALIHPCWVLTAAHCTDINTK 326
Qy 370 HLKVVLDQDLKEPFHQSFRVEKIFKYSYHNERDEIPHNDIALLLKPKVDGHCALLESK 429
Db 327 HLKVVLDQDLKEPFHQSFRVEKIFKYSYHNERDEIPHNDIALLLKPKVDGHCALLES 386
Qy 430 VYKTVCLPDGSPSPSSECHISGWVTETGKSRQLDADKVKLIANTLCSNRLQYDHMIDD 489
Db 387 VYKTVCLPDGSPSPSSECHISGWVTETGKSRQLDADKVKLIANTLCSNRLQYDHMIDD 446
Qy 490 SMICAGNLQKPDQCGDSGGLTCEKDGTYVYVYGVISWGLECGKRGVYVTVTKFLNW 549
Db 447 SMICAGNLQKPDQCGDSGGLTCEKDGTYVYVYGVISWGLECGKRGVYVTVTKFLNW 506
Qy 550 IKATIKSESG 559
Db 507 IKTVHREAG 516

RESULT 3
Q800Y7 PRELIMINARY; PRT; 540 AA.
ID Q800Y7 PRELIMINARY; PRT; 540 AA.
AC Q800Y7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hepatocyte growth factor activator (Fragment).
OS Meleagris gallopavo (Common turkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Meleagris.
OC NCBI_TaxID=9103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2212796; PubMed=12128063;
RA Holsberger D.R., Becker A.E., Thurston R.J., Rice C.D.;
RT "Expression of a hepatocyte growth-factor activator protein in turkey
RT (Meleagris gallopavo) deferent duct epithelial cells.";
RL Comp. Biochem. Physiol. 132:769-777(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Holsberger D.R., Becker A.E., Thurston R.J., Rice C.D.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY216598; AAC46038.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity and peptidolysis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR00742; EGF_2.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000083; Fibinctnl.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00058; FN1; 1.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
```

```
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS50070; KRINGLE_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
FT NON_TER 1 540
FT NON_TER 540 540
SQ SEQUENCE 540 AA; 61159 MW; 0BB3B4A89C0B577F CRC64;

Query Match 26.3%; Score 831; DB 13; Length 540;
Best local similarity 34.8%; Pred. No. 7.4e-68;
Matches 188; Conservative 86; Mismatches 208; Indels 58; Gaps 22;

Qy 47 EDYNOSENTSSTTHAENPD--WY- ----TEDQADQCPNCPCEHGGDC-LVHG-STFTCS 98
Db 21 EGRKREKDLCSYTHNYDRDKWGYCTLLSTSHSDYCANNPCQNGTCTFLAHRRTYHCA 80
Qy 99 CLAPFSGKCKQYONTCKONPC-----GRGCLITQSPPYRVCVKHPYTPSPCS 148
Db 81 CPBEFTGRDQ--MKKCFDDSLYEFDVDMWRSR-----VKQGSVEQCICVDGQM--ECL 131
Qy 149 QVV-PVCRPNPCGATCSRHKRSKFTCAQDPQKGFCEI--GSDDCVYDGYSGYRGKM 206
Db 132 RVEHKSCVHDPCWNGGECKMIASGKTVCDCKGPFVGVKYNIVPNHHCYRGNGTEYRGA 191
Qy 207 NRTVNOACLYMNSHLILQENYNNMEDAETHGIGEHFPCRNPDADKPCWCFIKVINDKV 266
Db 192 KTIISGHSCIPMNSDLLRELHVDSVEKAYOLGLGPFYCRNDEDEKPCYI-MKDNSL 250
Qy 267 KWECYDVSAQAQVAYP--ESPTPESTKLPGFDSGCKTEIAERKIK-RIYGGFKSTAG 323
Db 251 SWECYNTISCASRRRPVLEIDTFAVPRRP---CGRHKRKRFSVRPRLIGSSSLPG 306
Qy 324 KHPQASLQSSLPITISMPQHFQCGALIHPCWVLTAAHC--TDIKTRHLKVVLDGQDLK 381
Db 307 SHPWTAIIYG-----ESPCAGTLIQTQWVSAAHCFANSPKSIKVVLGQHFN 357
Qy 382 KEEFHEGSFVVEKIFKYSYHNERDEIPHNDIALLLKPKVDGHCALLESKYVTKLPDGS- 440
Db 358 RTDVTQTFEIKYILYQYVSRPTEH-DIALIKKNGQRCVAVKSQFVQPICLPESNT 416
Qy 441 -FPSGSECHISGWVTE---TGKGSROLLDAKVKLIANTLCSNRLQYDHMIDDSMICAGN 496
Db 417 VPDQFKQISGWHKHNITGY-SDVLOETLPIPEEKCRSPETIYGTISENMFCAGY 475
Qy 497 LQKPGQDTQCGDSGGLTCEKDGTYVYVYGVISWGLECGK--RPGVYTVTKFLNWKATI 554
Db 476 FDSK-SDACQDSGGGLACENNEISYLVISWGDGCGRVNKPQVYTRVPYVNWNERI 534

RESULT 4
Q8VCS4 PRELIMINARY; PRT; 653 AA.
AC Q8VCS4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; BC019376; AAH19376.1; -.
DR HSSP; P00761; IANI.
DR GO; GO:0005576; C:extracellular; IEA.
```

```
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR000083; Fibronectin.
DR InterPro; IPR000562; FN_Type_I.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001124; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00039; FN1.
DR Pfam; PF00040; FN2.
DR Pfam; PF00051; Kringle.
DR Pfam; PF00089; trypsin.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00113; FIBRONECTIN.
DR PRINTS; PR00018; KRINGLE.
DR PRODOM; PD000995; FN_Type_II.
DR PRODOM; PD000395; Kringle.
DR SMART; SM00181; EGF.
DR SMART; SM00059; FN2.
DR SMART; SM00130; KR.
DR SMART; SM00020; Tryp_Spc.
DR PROSITE; PS00022; EGF_1.
DR PROSITE; PS01186; EGF_2.
DR PROSITE; PS01233; FIBRONECTIN_1.
DR PROSITE; PS00023; FIBRONECTIN_2.
DR PROSITE; PS00021; KRINGLE_1.
DR PROSITE; PS00070; KRINGLE_2.
DR PROSITE; PS00240; TRYPSIN_DOM.
DR PROSITE; PS00134; TRYPSIN_HIS.
DR PROSITE; PS00135; TRYPSIN_SER.
KW Hypothetical protein; EGF-like domain; Glycoprotein; Hydrolase;
KW Kringle; Protease; Serine protease.
SQ SEQUENCE 653 AA; 70553 MW; FE18D90174ED6FDD CRC64;

Query Match 26.3%; Score 829; DB 11; Length 653;
Best Local Similarity 34.4%; Pred. No. 1.4e-67;
Matches 190; Conservative 79; Mismatches 187; Indels 96; Gaps 22;

QY 60 THAENPP--WYTEDQA-----DPCPNCEHGGDLV---HGSTFTCSCLAPRSG 105
DB 133 THNYDRDRAGVCAEVLTPVEGPAILDPCASGCLNGGTCSSTHHDGSG-YHSCSPLAFTG 191
QY 106 NKCKQVQNTCKD-----NPGCRGQCLITQSPPPYRCVKCKEPTG 144
DB 192 KDCG--TEKCFDETRYEVFEVGDHWARVSEGHVEQCQ---OMEGQA-----RCEDTTH--- 239
QY 145 PSCQVTVCPNPQNGATCSRHKRSKFTACPDQFKGFCEI-GSDCYVGDGYSYR 203
DB 240 -----TACLSSPCLNGGTCCHLVGTGTSVCTPLGYAGRFNCIVPTEHCFLGNGTEYR 292
QY 204 GKAKRTVNOHACLWNHLLQENYNNFMEDAEHGIQENFNCNPDADKPKCFIKVTN 263
DB 293 GVASTAASGLSCLAWNSDLLVQELHVDVSAVAALVGLGPHAYCRNPKDKDERPCYV-VKD 351
QY 264 DKVKEVCDVSAQAQD-----VAYPEESPTPESTKPLGFDSCGTEAERKIK- 312
DB 352 NALSWCYRLTACFESLARVHSQSPEILAALPESAPVRPT-----CGRHKRKTFLRP 404
QY 313 RIYGFKSTAGKHQWASLQSLPLTISMPQGHCGGALHPCWLTAAHC--TDIKTRH 370
DB 405 RIIGSSSLPFSHPWLAIIYG-----NSFCAGSLVHTCWVSAACHFANSPPRDS 455
QY 371 LKVLGDQDLKEEFHESQFVEKIFKYSHYNERDEIPHN-DIALKLKPVGDGHALESK 429
DB 456 ITVVLGQHFFNRITDVTQFGIEKYVPYTLYSVEN--PNNHDLVLRLKKGKGERCAVRS 513
QY 430 YKTVCLPD--GSFPGSGECHISGVGTE--TGKRSRQLLDAKVKLIANTLCLNSRQLYDH 485
```

```
514 FVQPICLPEAGSSFPFGHKQIAGWGHMBENVSSYNSLLEALVPLVADHKCSSPEVYGA 573
486 MIDDSMICAGNLQKPSQDTCQDSGSLTCEKDGTYVYVGIYSWGLECGK--RPGVYTOV 543
574 DISPNMLCAGYFDCK-SDACQDSGGPLVCEKRGVAYLYGLISWGGCGRLNKPGVYTRV 632
544 TKFLNWKATIK 555
533 ANYVDWINDRIR 644

RESULT 5
Q9BU99 PRELIMINARY; PRT; 516 AA.
AC Q9BU99;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to plasminogen activator, tissue.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC TISSUE=Skin;
RC SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR EMBL; BC002795; AA02795.1; -.
DR HSSP; P00750; 1A5H
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF_1.
DR Pfam; PF00051; Kringle; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR PRODOM; PD000395; Kringle; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00130; KR; 2.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00021; KRINGLE_1; 2.
DR PROSITE; PS00070; KRINGLE_2; 2.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
SQ SEQUENCE 516 AA; 57370 MW; BAB31901FDC96800 CRC64;

Query Match 24.3%; Score 766.5; DB 4; Length 516;
Best Local Similarity 35.0%; Pred. No. 6.2e-62;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFGSGNKCO-KVQNTCKNKP--CGRG----- 123
DB 40 CSEPCFNGGTCCQALYFSDFCVQCPEGFAGKCEIDTRATCYEDQGSIYGTWSTABSG 99
QY 124 -QLITQSPPPYRCVKCKEPTGTSQSVVPCVRPNPCQ-----NGATCSRHKRSKFTCAC 178
```

Db 100 AECTNWS-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDNRDPSKPCYV 146  
Qy 179 PDQFK-GKF-----CEIGSDCYVDGYSYRGKMRNTVNOHACLWNHLLQENY 228  
Db 147 ---FKAGYSSEPCSTPACSEGNDCYFGNGSAYRGTHSLTESGASCLPWNMILIGVY 203  
Qy 229 NMFEDAETHGIGEHNFNCRNPDADKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEESP 288  
Db 204 TAQNPQAALGLGKHNYCRNPDGAKPCHV-LKRRLLTWELWYCDVPSCS-----251  
Qy 289 TEPSTKLPGFDSGCKTEIAERKIKIYGGFKSTAGKHPWQASLQSLPLTISMPQGHFCG 348  
Db 252 -----TCGLRQYSQPF-RIKGLFADIASHPWQAIAFAK--HRRSPGERFLCG 297  
Qy 349 GALIHCPCWLTAAHCTD--IKTRHLKVLGDQDLKKEEFHESRVEKIFKYSHYNERDE 406  
Db 298 GILISSCWILSAHCFQBRFPFHLLTVILGTYRVVPEGEERQKFEVEKIVHKEFD--DD 355  
Qy 407 IPHNDIALKLKPDVGHCALESKYKVTCLP--DGSFPGSGECHISGWTETGKG--SR 462  
Db 356 TYNDNIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTCELSGVGKHEALSPFYSE 415  
Qy 463 QLLDAKVKLIANTLNSRQLYDHMDSDMICAGNLQKPG-----QDTCQDSDGGLPCTCK 517  
Db 416 RLKEAHVRLYPSRRTSQHLLNRTVTDNMLCAGDTRSGGPOANLHDACQDSDGGLPVLCLN 475  
Qy 518 DGTYYVYGVISWGLECGKR--PGVYTVQTKFLNWKATIK 555  
Db 476 DGRMTLVGLIISWGLCGQKDPGVYTKVTNYLDWIRDNR 515  
RESULT 6  
Q86YK8  
ID Q86YK8 PRELIMINARY; PRT; 562 AA.  
AC Q86YK8  
DT 01-JUN-2003 (Tremblrel. 24, Created)  
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Tissue plasminogen activator.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Liu Y., Xu L., Zeng Y., He X.;  
RT "cDNA of tissue plasminogen activator."  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY221101; RAO34406.1; --  
DR GO; GO:0005576; Cextracellular; IEA.  
DR GO; GO:0004263; Fichymotrypsin activity; IEA.  
DR GO; GO:0004295; Fitypsin activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR009003; Cys Ser trypsin.  
DR InterPro; IPR006209; EGF like.  
DR InterPro; IPR000083; Fibrinctnl.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00008; EGF; 1.  
DR Pfam; PF00039; fnl; 1.  
DR Pfam; PF00051; kringle; 2.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PRINTS; PR00013; KRINGLE.  
DR ProDom; PD000395; Kringle; 2.  
DR SMART; SM00058; FN1; 1.  
DR SMART; SM00130; KR; 2.  
DR SMART; SM00020; Tryp SPC; 1.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01253; FIBRONECTIN\_1; 1.  
DR PROSITE; PS00021; KRINGLE\_1; 2.

DR PROSITE; PS50070; KRINGLE\_2; 2.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
SQ SEQUENCE 562 AA; 62902 MW; 837D98392F6EDD1F CRC64;  
Query Match 24.3%; Score 765.5; DB 4; Length 562;  
Best Local Similarity 35.0%; Pred. No. 8.6e-62;  
Matches 182; Conservative 76; Mismatches 177; Indels 85; Gaps 20;  
Qy 77 CQPNCEHGGDC--LVHGSTFTCSCLAPSPGNKCO-KVQNTCKDNP--CGRG-----123  
Db 86 CSEPRCFNGTCCQALYFSDFCQCFEPGAGKCCIDTRATCYEDQGISYRGTWTAESG 145  
Qy 124 -OCLITQSPPYKRCVKCHPYTGTPSCSQVVPVCRPNPCQ-----NGATCSRHRKRSKFTCAC 178  
Db 146 AECTNWS-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDNRDPSKPCYV 192  
Qy 179 PDQFK-GKF-----CEIGSDCYVDGYSYRGKMRNTVNOHACLWNHLLQENY 228  
Db 193 ---FKAGYSSEPCSTPACSEGNDCYFGNGSAYRGTHSLTESGASCLPWNMILIGVY 249  
Qy 229 NMFEDAETHGIGEHNFNCRNPDADKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEESP 288  
Db 250 TAQNPQAALGLGKHNYCRNPDGAKPCHV-LKRRLLTWELWYCDVPSCS-----297  
Qy 289 TEPSTKLPGFDSGCKTEIAERKIKIYGGFKSTAGKHPWQASLQSLPLTISMPQGHFCG 348  
Db 298 -----TCGLRQYSQPF-RIKGLFADIASHPWQAIAFAK--HRRSPGERFLCG 343  
Qy 349 GALIHCPCWLTAAHCTD--IKTRHLKVLGDQDLKKEEFHESRVEKIFKYSHYNERDE 406  
Db 344 GILISSCWILSAHCFQBRFPFHLLTVILGTYRVVPEGEERQKFEVEKIVHKEFD--DD 401  
Qy 407 IPHNDIALKLKPDVGHCALESKYKVTCLP--DGSFPGSGECHISGWTETGKG--SR 462  
Db 402 TYNDNIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTCELSGVGKHEALSPFYSE 461  
Qy 463 QLLDAKVKLIANTLNSRQLYDHMDSDMICAGNLQKPG-----QDTCQDSDGGLPCTCK 517  
Db 462 RLKEAHVRLYPSRRTSQHLLNRTVTDNMLCAGDTRSGGPOANLHDACQDSDGGLPVLCLN 521  
Qy 518 DGTYYVYGVISWGLECGKR--PGVYTVQTKFLNWKATIK 555  
Db 522 DGRMTLVGLIISWGLCGQKDPGVYTKVTNYLDWIRDNR 561  
RESULT 7  
Q8SQ23  
ID Q8SQ23 PRELIMINARY; PRT; 562 AA.  
AC Q8SQ23  
DT 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE T-plasminogen activator.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA TISSUE=Enamel organ;  
RA Ding Y., Xue J., Bartlett J.D.;  
RT "T-plasminogen activator in tooth tissues."  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.  
DR EMBL; AF364605; AAM00297.1; --  
DR HSRF; P00761; IAN1.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004295; F:trypsin activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.



```

DR InterPro: IPR009003; Cys_Ser trypsin.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR000083; Fibrinectn1.
DR InterPro: IPR006210; IEFG.
DR InterPro: IPR000001; Kringle.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1A.
DR Pfam: PF00008; EGF_1.
DR Pfam: PF00039; fnl; 1.
DR Pfam: PF00051; kringle; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRINTS: PR00018; KRINGLE.
DR PRODOM: PD000395; Kringle; 2.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00058; FNI; 1.
DR SMART: SM00130; KR; 2.
DR SMART: SM00130; KR; 2.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01253; FIBRONECTIN_1; 1.
DR PROSITE: PS00021; KRINGLE_1; 2.
DR PROSITE: PS00070; KRINGLE_2; 2.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
KW
SQ
Query Match 24.0%; Score 758.5; DB 6; Length 562;
Best Local Similarity 33.5%; Pred. No. 3.8e-61;
Matches 183; Conservative 80; Mismatches 201; Indels 83; Gaps 19;

QY 49 YNQEENTSSLTTHAENPDWYTDQDPCQPNP-----CEHGDCGL--VHGSFTCS 98
DB 50 YQHCQWLRLPRLNRVHEHCWCDGQTCQSHVPKSCBPCFNGGTCLQAIYFSDVCQ 109
QY 99 CLAFPSGNKQKQ-KVQNTCKNCP-CGRGCLITQSPPYRCV-----CKHPTGPS 148
DB 110 CPVGFGRQCEIDARATCYEDQGITRGTTSTTESGA--ECVNNTSGLASMEYNGR--- 164
QY 149 QVVPVCPNFCQ-----NGATCSRHKRSKFTC-----ACDPQFKGCEIGSDCYNG 197
DB 165 -----RFDVAVKLGNGHNYCRNPDKDKSPWCYIFKAEKYSDFPCSTPACTKEKCYTG 218
QY 198 DGSYRGKRNRTVQHAACLYWNSHLLQENYMFEDAEHGHGECNFCRNPDADEKPMC 257
DB 219 KGLDYGRTRSLTMSGATCLPWSIVLAKIYTAWSNAQTLGLGKHNYCRNPDGDTQPCW 278
QY 258 FIKTNDKVKWEYCDVSACSNQDVAYEESFTEPSTKLPGFDCGKTEIAERKILYGG 317
DB 279 HV-LKDHKLWYCDLPQCV-----TCGURQYKEPQF-RIKGG 314
QY 318 FKSTAGKHPMAQSLQSLPLTISMPQGHFCGALHPWCVLTAHCTD--IKTRHLKVL 375
DB 315 LYADITSHPWAAI--FVKNRSRGERFLCGGILISSCWLSAAHCFQERFPHVVRVL 372
QY 376 GDQDLKEEFHQFRVEKIFKYSHYNERDIPNDIALKLKVPDGHCALESKYVKTVC 435
DB 373 GRVRLVPEGEQAFVEKIVYHKEFD--DDTYNDIALQLKSDSLTCAQESDAVRTVC 430
QY 436 LPDGS--PPSGSECHISGWYTERGKG--SRQLLDKVKLIANTLCNSRQLYDHMIDDSM 491
DB 431 LPEANLQLPDWTCELSGYGKEASSPPYSERLKEAHVRLYPSRCLTKHFNKTIINM 490
QY 492 ICAGNLQKPG-----QDTQCGSDSGPLTCEKDGKDYVYVGVISWGLECGKR--PGVYQV 544
DB 491 LCAGTRSGDNANLHDAQCQSDSGPLVCMKGNEMTLVGVISWGLECGKDYVYVKTVC 550
QY 545 KFLANWK 551
DB 551 NYLANWR 557

```

```

RESULT 8
Q8MKXB1
ID Q8MKXB1 PRELIMINARY; PRT; 564 AA.
AC Q8MKXB1
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Tissue-type plasminogen activator.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_taxid=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Sugiki M., Yoshida E., Anai K., Maruyama M.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: CONTAINS 2 KRINGLE DOMAINS.
DR EMBL; AY029518; AAK40240.1;
DR GO; GO:0005578; C:extracellular; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0006233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR009003; Cys_Ser trypsin.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR00083; Fibrinectn1.
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR000001; Kringle.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001314; Peptidase_S1A.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF00039; fnl; 1.
DR Pfam: PF00051; kringle; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR PRODOM: PD000395; Kringle; 2.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00058; FNI; 1.
DR SMART: SM00130; KR; 2.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01253; FIBRONECTIN_1; 1.
DR PROSITE: PS00021; KRINGLE_1; 2.
DR PROSITE: PS00070; KRINGLE_2; 2.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;
KW Serine protease.
KW
SQ
Query Match 23.4%; Score 737; DB 6; Length 564;
Best Local Similarity 35.5%; Pred. No. 3.7e-59;
Matches 182; Conservative 73; Mismatches 180; Indels 78; Gaps 21;

QY 77 CQNPCEHGGDC--LVHGSTTCCLAPFSGNKQKQ-KVQNTC-KDNPCG-EGQCLITOSP 131
DB 87 CSEPCRLNGTCSQALYFSDFCQCFEGVGRKCEVDTRACVYEDRGIGYGTWSTTESG 146
QY 132 PYRVC-----CHPYTGFSCSQVPEVCRPNFCQ-----NGATCSRHKRSKFTCAC-- 178
DB 147 A--QCVNWNSSWLAKPYSGR-----KFNALRLGNGHNYCRNPDRTKPCYVFR 195
QY 179 PQDFKGFCEI-----GSDCCYVGDGYSYRKKRNRTVQHAACLYWNSHLLQENYMF 232
DB 196 AGTYSPEFCSTPACSKENKNGCYLGKQAYRGTHTSLTSGASCLFPWNSMLLVGEYTRQ 255
QY 233 EDAETHGIGEHNFCRNPDADEKPFVKVTNDKVKWEYCDVSACSAQDVAYPESSPTPS 292

```





Query Match		22.6%; Score 712; DB 4; Length 395;
Best Local Similarity		35.1%; Pred. No. 4.7e-57;
Matches 162; Conservative		67; Mismatches 143; Indels 90; Gaps 15;
QY	114	TKDNPCGRGCLITQSPYRCVKHYTGSPSCQVVPORNP-----CQNG-ATCSR 167
DB	3	SCRDE---KTMIVQHQSWLR-----PVLRSRVEYCNCSGRAQCS 43
QY	168	HKRSKFTCACDQFKGFCETGSDCYVGGYSYRGHNRVTHQACLYWNHLLQEN 227
DB	44	-----GNSDCYFGNGSAVGRGTHSITESGASCLPWNMILICKV 81
QY	228	YNMFMEDATHGIGHNFNCPNDAEKPCFKITVNDKVKMEYCDVSACSADVAPEES 287
DB	82	YTAQNPSCAALGLGKHNYCRNPDGAKPWCHV-LKNRLTWEYCDVPSCS----- 130
QY	288	PTEPTKLPFGSCGKTHIAERKIRIYGGFKSTAGKHPQASLQSSPLTIISMPQGHF- 346
DB	131	-----TCGLRQYSQGF-RIKGGLFADTASHPWQDAIPAKHE---ESPGERFL 174
QY	347	CGGALIHPCWLVTAHCTD--IKRHLKVVLDGDLKKEEFHQFRVEKIFKYSYHNER 404
DB	175	CGGILISSCWILSAACFOERPPHLLTVILGRYRVVPGEBEQFCEKIVHKEFD-- 232
QY	405	DEIPHNDIALKLKVDGHCALSKYKTVCLP--DGSFFPSGSECHISGMGVETGKG-- 460
DB	233	DDTYNDIALLOKSDSSCAQESSVVRTVCLPPADLQLPDMTECELSYGKHEALSIFY 292
QY	461	SQLLDKAVKLANTLCNSRQYLDHMDSMICAGNLQKPG-----QDTCQDSDGGLTC 515
DB	293	SERLKEAHVRLPSSRCTSQHLLNRITVDMLCAGTRSGGQANLHDACQDSDGGLVC 352
QY	516	EKDGTYYVGVISWGLECKR--PGVYTVTKFLNWKATIK 555
DB	353	LNDGRMTLVGIISWGLGCGCKQDVPVYTKVTNYLDWIRDNR 394
RESULT 12		
ID	O35727	PRELIMINARY; PRT; 597 AA.
AC	O35727;	
DT	01-JAN-1998	(T-EMBLrel. 05, Created)
DT	01-JAN-1998	(T-EMBLrel. 05, Last sequence update)
DT	01-OCT-2003	(T-EMBLrel. 25, Last annotation update)
DE	Factor XII.	
GN	F12.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Liver;	
RA	Schloesser M., Schwager S., Engel W.,	
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.	
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.	
CC	-1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.	
DR	EMBL; X99571; CAA67891.1; --	
DR	HSSP; P00760; 1AQ7.	
DR	MEKOPS; S01.211; --	
DR	MGD; MGI:1891012; F12.	
DR	GO; GO:0005576; C:extracellular; IEA.	
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.	
DR	GO; GO:0008233; F:peptidase activity; IEA.	
DR	GO; GO:0004295; F:trypsin activity; IEA.	
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.	
DR	InterPro; IPR009003; Cys Ser trypsin.	
DR	InterPro; IPR006209; EGF-like.	
DR	InterPro; IPR000083; Fibnctnl.	
DR	InterPro; IPR000562; FN Type II.	
DR	InterPro; IPR006210; IEGF.	
DR	InterPro; IPR000001; Kringle.	
DR	InterPro; IPR001254; Peptidase_S1.	
InterPro; IPR001314; Peptidase_S1A.		
Pfam; PF00008; EGF; 2.		
Pfam; PF00039; fn1; 1.		
Pfam; PF00040; fn2; 1.		
Pfam; PF00051; kringle; 1.		
Pfam; PF00083; trypsin; 1.		
PRINTS; PRO0722; CHYMOTRYPSIN.		
PRINTS; PRO0013; FNTYPII.		
PRINTS; PD00018; KRINGLE.		
ProDom; PD000995; FN Type II; 1.		
ProDom; PD000395; Kringle; 1.		
SMART; SM00181; EGF; 2.		
SMART; SM00058; FN1; 1.		
SMART; SM00059; FN2; 1.		
SMART; SM00130; KR; 1.		
SMART; SM00020; Tryp_SPC; 1.		
PROSITE; PS00022; EGF 1; 2.		
PROSITE; PS01186; EGF 2; 1.		
PROSITE; PS01253; FIBRONECTIN 1; 1.		
PROSITE; PS00023; FIBRONECTIN 2; 1.		
PROSITE; PS00021; KRINGLE 1; 1.		
PROSITE; PS00070; KRINGLE 2; 1.		
PROSITE; PS00240; TRYPSIN DOM; 1.		
PROSITE; PS00134; TRYPSIN HIS; 1.		
PROSITE; PS00135; TRYPSIN SER; 1.		
EGF-like domain; Glycoprotein; Hydrolase; Kringle; Protease;		
Serine protease.		
KW		
Sequence		597 AA; 65638 MW; P3AC07C37D0C0PBA CRC64;
Query Match		22.0%; Score 694; DB 11; Length 597;
Best Local Similarity		32.8%; Pred. No. 3.7e-55;
Matches 189; Conservative		69; Mismatches 191; Indels 127; Gaps 29;
QY	56	SSILTHAENPDWY--TEDQAPC-OPNCEHGGDCLVHGSTFTCSCLAPFGNKCQKV 111
DB	74	ATTENFDENQWGYCLEPFKKVKDKCHKNPCHKGTCINTPNGPHCLCPHLTGKCHQK- 132
QY	112	QNTCKDNPCGRGCLITQSPYY-----RCVCK--HPYTGSCSVQVPVC 154
DB	133	-----EKCFEPOLLKFFHENELWFKTGPGGVARCEKSGSEAHCKPVASQ---AC 178
QY	155	RPNCPQNGATC---SRHKRSKTCACPDQFKGKCEIGS-DCVYVDGYSYRGKMRV 210
DB	179	SINPLNGSGCLLVEDHP-----LCRCPTGYTGFCDDLWATCYEGRGUSYRGAGTTQ 233
QY	211	NOHACLYWNHLLQENY-NMFMEDATHGIGHNFNCPNDAEKPCFKITVNDKVKWE 269
DB	234	SGAPCQRT---VEATRYNWKTEQALSWGLGHAFCRNPDNDRPWCFFV-WSGDRLSWD 288
QY	270	YCDVSACSADVA-----YPSPTPE-PSKLPGPS-----CGKTEIAE 308
DB	289	YCGLEQCQTPTTFAPLVVPESQESPSQAPSLSHAPNDSTDHQTSLSTNTMGCGQ---RP 345
QY	309	RK---IKRIYGGFKSTAGKHPQASLQSSPLTIISMPQGHFCGGLIHPCWLVTAHCT 364
DB	346	RKGLSSFWRVVGLVALPGSHFYAALYWG-----NNFCAGSLIAFCWLVTAHCL 396
QY	365	DIK--TRHLKVVLDGDLKKEEFHQ-----SFRVEKIFKYSYHNERDEIHPND 411
DB	397	QNRPAPELITVLG-----QDRHNQSCWQCOTLAVRSYRLHEGFSSITY-----QHD 443
QY	412	IALLKL-KPVDGHCALSKYKTVKCLDGGSPSGSE--CHISGMGVETG--KGSROLLD 466
DB	444	LALLRQESKTNSCAILSPHVQVPCLPESGAAPSETVLCVAGHGHQLEGAEFYSTLOE 503
QY	467	ARKVLIANTLCNSRQYLDHMDSMICAGNLQKPGQDTCQDSDGGLPTCKEDGT-----YY 522
DB	504	AQVPFIALLDRCSNHNHGDALLPQMLCAGFLE-GTDAQCQDSDGGLVCB-EGTAERHLT 561
QY	523	VGVISWGLECKR--PGVYTVTKFLNWKATIKS 556
DB	562	LRGVISWSGCGDRNKPVGVTVDVANYLAWIQKHAS 597

QY	155	RPNFCNGATC---SHKRRSFTCA	CPDQFQKGFCEIGS-DDCYVGDGYSGEKKNRVT	210
Db	191	SINPCLNGSSCLVADHP-----LCRCPTGYTG	YFCDLDMATWATYEGRGLSYHGQAGTTQ	245
QY	211	NQHACLYVNSHLLQENY-NMFEDEATETIGIS	HNFRCNPDADAEKPCWCFIKVTNDKVKWE	269
Db	246	SGAPCQRTW-----VEATYRNWTEKQALS	WGLGHHAFCRNPNDNTRPWCFF-WSGDRLSWD	300
QY	270	YCDVVSACSAQDVA-----YPRESPTB	-PSTKLPGFDS-----CGKTEIAE	308
Db	301	YCGLEQCQTTFAPLVVPESQESPSQAPSL	SHAPNDSTDHQTSLSKNTMTMGCGQ----	357
QY	309	RK-----IKRTYGFKSTAGKHPQASLOS	SLPLTISMPQGHFCGGLIHPFCWLTAAHCT	364
Db	358	RKGLSSPMRVVGGVLVALPGSHPIALYWG	-----NNFCAGSLIAPCWLTAAHCL	408
QY	365	DIK--TRHLKVLGDQDLKKKEPHEQ-----	SFVEKIPKYSHYNERDEIHPND	411
Db	409	QNRPAPEELTVILG-----QDRHNSCEWC	QTLAVRSYRLHGFSSITY-----OHD	455
QY	412	IALLKL-KPVDGHCALESKVYKVTVC	LPDGSPPSGSE--CHISGMGVTETG--KGSROL	466
Db	456	LALLRQESKTNSCAIIISHPVQVCLPSGA	APSPETVLCEVAGHGHQFEGAAEEYSTLQE	515
QY	467	AKVKLIANTLCNSQLYDHMIDDSMICAGN	LQKPGQDTCQDSDGGPITCSKDGTF-----	522
Db	516	AQVFFIALDCGNSNVHGDAILPCLM	CAGFLE-GGTDACQDSDGGPLVCB-EGTAEH	573
QY	523	YGVTVSWGLECGKR--PGVYQVTKFLN	MIKATIKS	556
Db	574	LRGVISWGGCGDRNKPFGVTDVANYL	AWIQKHAS	609

RESULT 14

Q8MILO

PRELIMINARY; PRT; 433 AA.

Q8MILO;

Q1-OCT-2002 (TReMBLrel. 22, Created)

Q1-OCT-2002 (TReMBLrel. 22, Last sequence update)

Q1-OCT-2003 (TReMBLrel. 25, Last annotation update)

Urokinase-type plasminogen activator.

PLAU.

OS Oryctolagus cuniculus (Rabbit)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI\_TaxID=9986;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22155945; PubMed=12149463;

RA Falkenberg M., Tom C., DeYoung M.B., Wen S., Linnemann R., Dichek D.A.;

RA "Increased expression of urokinase during atherosclerotic lesion development causes arterial constriction and lumen loss, and accelerates lesion growth."

RT Proc. Natl. Acad. Sci. U.S.A. 99:10665-10670(2002).

RL CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.

CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.

DR EMBL; AY122285; AA083187.1; -.

DR GO; GO:0004263; F:chymotrypsin activity; IEA.

DR GO; GO:0016301; F:kinase activity; IEA.

DR GO; GO:0008233; F:peptidase activity; IEA.

DR GO; GO:004295; F:trypsin activity; IEA.

DR GO; GO:0004508; F:proteolysis and peptidolysis; IEA.

DR InterPro; IPR009003; Cys\_Ser\_trypsin.

DR InterPro; IPR006209; EGF-like.

DR InterPro; IPR000001; Kringle.

DR InterPro; IPR001254; Peptidase S1.

DR InterPro; IPR001314; Peptidase S1A.

DR InterPro; IPR008293; Pept\_S1A\_UPA.

DR Pfam; PF00051; kringle; 1.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PRINTS; PR00019; KRINGLE.

```

RESULT 13
Q80YCS PRELIMINARY; PRT; 609 AA.
ID Q80YCS;
AC Q80YCS;
DT 01-JUN-2003 (trEMBLrel. 24, Created)
DT 01-JUN-2003 (trEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (trEMBLrel. 25, Last annotation update)
DE Similar to coagulation factor XII (Hageman factor) (Fragment).
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_Taxid=10090;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=FVB/N; TISSUE=Liver;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC049867; AAH49867.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser_trypsin.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000083; Fibrinctnl.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00039; fn1; 1.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00013; FNTYPEII.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000995; FN_Type_II; 1.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00058; FN1; 1.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00130; KR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01253; FIBRONECTIN_1; 1.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00020; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
FT NON_TER
SQ SEQUENCE 609 AA; 66793 MW; DF97D4DB2369B6D2 CRC64;

Query Match 21.8%; Score 689; DB 11; Length 609;
Best Local Similarity 32.6%; Pred. No. 1.1e-54;
Matches 188; Conservative 70; Mismatches 191; Indels 127; Gaps 29

Qy 56 SSTLTAAHPDWY---TEDQADPC-QNPCEHGGDCLVHGSTFTCSCLAPFSGNKKCVK 111
Db 86 ATTTFNFDQQQYQYCLEPKVKXDKHCKNFCCHKGTCTINTFNGPCLCPHELTGKHCQK- 144
Qy 112 QNTCKNDPCGRGQCLITSPYY-----RCVCK--HPYTGPCSQVWPVC 154
Db 145 -----EKCFEPOLLKFFHFNELWFRTPGGGVARCECKGSEAHCKPVASQ---AC 190

```

```

DR ProDom; P0000395; Kringle; 1.
DR SMART; SMO0130; KR; 1.
DR PROSITE; PS00020; TRYP_SPC; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS00021; KRINGLE_2; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYP_SIN_DOM; 1.
DR PROSITE; PS00134; TRYP_SIN_HIS; 1.
DR PROSITE; PS00135; TRYP_SIN_SER; 1.
DR PIRSF; PIRSF001144; Utk_plasm_act; 1.
KW Glycoprotein; Hydrolase; Kinase; Kringle; Protease; Serine protease.
SQ SEQUENCE 433 AA; 48375 MW; 65E64F36415549B0 CRC64;

Query Match
Best Local Similarity 21.2%; Score 667.5; DB 6; Length 433;
Matches 155; Conservative 62; Mismatches 163; Indels 37; Gaps 14;

QY 159 CONGATCSRHRKRSK-FTCACPQPKGKFCFEGS-DDCYVGDGYSYRGQMTVNHACL 216
DB 35 CLNGGTCVTYKFSNIWRCNCPKFKQGEHCIEDTLTKTCYHGDSYRGKANTDINDRPCL 94

QY 217 YNSHLLLOENYNMFWDAETHGIGHNFCRNPDADKPCFIKVTNDKWKVEYCDVSAC 276
DB 95 ANSANVLTKYTHARPDALQLGKHNYCRNPDHQRFPWCIVQGLKQLQGE-CKVHDC 153

QY 277 SAQDVAYPESPTPEPTKLPFGDSCKTEIAERKIRIYGGPKSTAGKHPWQASLQSSLP 336
DB 154 SS-----GKKPALPPGKLE--FQCQKALRPR--FKIIGGEFTIENQPWFAAI----- 198

QY 337 LTISMPQG---HFCGGALHPCWVLTAAHC--TDIKTRHLKVLGDQDLKKEEFHQSF 390
DB 199 --YRHRGSGVTYVCGSLISPCWVVSATHCFINHQKEDYIVYLGSRSLNMTPEMKF 256

QY 391 RVEKIFKYSYNERDEIPHNDIALKLPVDGHCALSKYKVTCLP----DGSPFSGSE 446
DB 257 EVEQLILHEGYRADTLAHHNDIALKILSNNGCQAQPSRSIQITICLPWNADPNF--GTS 314

QY 447 CHISGCVGTETGK--GSRQLLDKAVKLIANTLCNSRQLYDHMDSDMICAGNLQKPGDGT 504
DB 315 CEITGFGKENSTDYLYPEQLKMTVKLVSYQEQPHYGVSEVTTKMLCAADPQWE-TDS 373

QY 505 CQDSSGGLPCKEKGDTGYVYVGVSWGLECG--KRPGVYTVQTKFLNWKATIKSSSG 559
DB 374 CQDSSGGLPVCVQGRMTLTGIVSWGRGALKNKPGVTRVSRFLPWIRSHIGEENG 430

RESULT 15
Q8MHY7 PRELIMINARY; PRT; 433 AA.
AC Q8MHY7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Urokinase-type plasminogen activator.
GN UROKINASE.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Sugiki M., Yoshida E., Anai K., Maruyama M.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yano W., Watanabe M.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP "Oryctolagus cuniculus urokinase-type plasminogen activator, mRNA, complete cds."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.
DR EMBL; AY029517; AAK40239.1; -.
DR EMBL; AB087224; BAC02685.1; -.

```

```

DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR006203; EGF-like.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001314; Peptidase S1A.
DR InterPro; IPR008293; Pept S1A_UPA.
DR Pfam; PF00051; kringle; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR0018; KRINGLE.
DR ProDom; PD000395; Kringle; 1.
DR SMART; SMO0130; KR; 1.
DR SMART; SMO0020; TRYP_SPC; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS00021; KRINGLE_1; 1.
DR PROSITE; PS00070; KRINGLE_2; 1.
DR PROSITE; PS00240; TRYP_SIN_DOM; 1.
DR PROSITE; PS00134; TRYP_SIN_HIS; 1.
DR PROSITE; PS00135; TRYP_SIN_SER; 1.
DR PIRSF; PIRSF001144; Utk_plasm_act; 1.
KW Glycoprotein; Hydrolase; Kinase; Kringle; Protease; Serine protease.
SQ SEQUENCE 433 AA; 48444 MW; 6DD35A371010A6EE CRC64;

Query Match
Best Local Similarity 21.1%; Score 666.5; DB 6; Length 433;
Matches 155; Conservative 61; Mismatches 164; Indels 37; Gaps 14;

QY 159 CONGATCSRHRKRSK-FTCACPQPKGKFCFEGS-DDCYVGDGYSYRGQMTVNHACL 216
DB 35 CLNGGTCVTYKFSNIWRCNCPKFKQGEHCIEDTLTKTCYHGDSYRGKANTDINDRPCL 94

QY 217 YNSHLLLOENYNMFWDAETHGIGHNFCRNPDADKPCFIKVTNDKWKVEYCDVSAC 276
DB 95 ANSANVLTKYTHARPDALQLGKHNYCRNPDHQRFPWCIVQGLKQLQGE-CKVHDC 153

QY 277 SAQDVAYPESPTPEPTKLPFGDSCKTEIAERKIRIYGGPKSTAGKHPWQASLQSSLP 336
DB 154 SS-----GKKPALPPGKLE--FQCQKALRPR--FKIIGGEFTIENQPWFAAI----- 198

QY 337 LTISMPQG---HFCGGALHPCWVLTAAHC--TDIKTRHLKVLGDQDLKKEEFHQSF 390
DB 199 --YRHRGSGVTYVCGSLISPCWVVSATHCFINHQKEDYIVYLGSRSLNMTPEMKF 256

QY 391 RVEKIFKYSYNERDEIPHNDIALKLPVDGHCALSKYKVTCLP----DGSPFSGSE 446
DB 257 EVEQLILHEGYRADTLAHHNDIALKILSNNGCQAQPSRSIQITICLPWNADPNF--GTS 314

QY 447 CHISGCVGTETGK--GSRQLLDKAVKLIANTLCNSRQLYDHMDSDMICAGNLQKPGDGT 504
DB 315 CEITGFGKENSTDYLYPEQLKMTVKLVSYQEQPHYGVSEVTTKMLCAADPQWE-TDS 373

QY 505 CQDSSGGLPCKEKGDTGYVYVGVSWGLECG--KRPGVYTVQTKFLNWKATIKSSSG 559
DB 374 CQDSSGGLPVCVQGRMTLTGIVSWGRGALKNKPGVTRVSRFLPWIRSHIGEENG 430

Search completed: May 24, 2004, 09:48:56
Job time : 41 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 09:45:05 ; Search time 17.5 Seconds  
(without alignments)  
1652.031 Million cell updates/sec

Title: US-09-912-559-3  
Perfect score: 3154  
Sequence: 1 MFARMSDLHVLLMALVYKLT.....TQVTKFLNWKIKESGPF 560

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	832.5	26.4	655	1 US-08-148-910-12	Sequence 12, Appl
2	832.5	26.4	655	1 US-08-448-937A-12	Sequence 12, Appl
3	786	24.9	562	6 5244676-5	Patent No. 5244676
4	777.5	24.7	527	6 5520913-1	Patent No. 5520913
5	767.5	24.3	527	2 US-08-811-949-39	Sequence 39, Appl
6	766.5	24.3	527	1 US-07-609-510B-16	Sequence 16, Appl
7	766.5	24.3	527	5 PCT-US91-01025A-2	Sequence 2, Appl
8	766.5	24.3	527	6 5185259-8	Patent No. 5185259
9	766.5	24.3	562	2 US-08-811-949-43	Sequence 43, Appl
10	766.5	24.3	562	2 US-08-560-098A-50	Sequence 50, Appl
11	766.5	24.3	562	2 US-08-883-795A-38	Sequence 38, Appl
12	766.5	24.3	562	2 US-09-703-695A-4	Sequence 4, Appl
13	766.5	24.3	562	6 5185259-3	Patent No. 5185259
14	766.5	24.3	562	6 5200340-2	Patent No. 5200340
15	766.5	24.3	562	6 5344773-2	Patent No. 5344773
16	734.5	23.3	472	2 US-08-811-949-63	Sequence 63, Appl
17	731	23.2	477	2 US-08-560-098A-51	Sequence 51, Appl
18	728	23.1	437	2 US-08-811-949-51	Sequence 51, Appl
19	727	23.1	437	2 US-08-811-949-49	Sequence 49, Appl
20	722	22.9	437	2 US-08-811-949-57	Sequence 57, Appl
21	721	22.9	437	2 US-08-811-949-55	Sequence 55, Appl
22	720	22.8	546	6 5200340-6	Patent No. 5200340
23	715.5	22.7	378	4 US-09-553-498-10	Sequence 10, Appl
24	715.5	22.7	378	4 US-09-618-869-10	Sequence 10, Appl
25	710.5	22.5	355	2 US-08-811-949-47	Sequence 47, Appl
26	709.5	22.5	355	1 US-08-137-116-1	Sequence 1, Appl
27	709.5	22.5	355	1 US-08-217-618-1	Sequence 1, Appl

28	709.5	22.5	355	1 US-08-427-640-2	Sequence 2, Appl
29	709.5	22.5	355	1 US-08-217-617A-1	Sequence 1, Appl
30	709.5	22.5	355	1 US-08-217-616-1	Sequence 1, Appl
31	709.5	22.5	355	2 US-08-811-949-45	Sequence 45, Appl
32	709.5	22.5	355	3 US-08-794-528-1	Sequence 1, Appl
33	709.5	22.5	355	6 5223256-1	Patent No. 5223256
34	709	22.5	347	2 US-08-811-949-1	Sequence 1, Appl
35	707.5	22.4	355	1 US-08-427-640-6	Sequence 6, Appl
36	707.5	22.4	389	2 US-08-811-949-65	Sequence 65, Appl
37	706.5	22.4	355	2 US-08-811-949-59	Sequence 59, Appl
38	703.5	22.3	355	2 US-08-811-949-53	Sequence 53, Appl
39	702.5	22.3	389	2 US-08-811-949-67	Sequence 67, Appl
40	698	22.1	354	2 US-08-811-949-61	Sequence 61, Appl
41	698	22.1	356	1 US-08-427-640-4	Sequence 4, Appl
42	684.5	21.7	356	1 US-08-427-640-8	Sequence 8, Appl
43	663	21.0	430	6 5219569-2	Patent No. 5219569
44	662.5	21.0	411	3 US-09-181-816-1	Sequence 1, Appl
45	661.5	21.0	411	1 US-08-087-163-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-08-148-910-12  
; Sequence 12, Application US/08148910  
; Patent No. 5466593  
; GENERAL INFORMATION:  
; APPLICANT: Takeshi SHIMOMURA et al.  
; TITLE OF INVENTION: No. 5466593el Protein and Gene Encoding Said Protein  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch,  
; MEDIUM TYPE: 500 KB Storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/148,910  
; FILING DATE: No. 5466593ember 5, 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warren M. Cheek, Jr.  
; REGISTRATION NUMBER: 33,367  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-8850  
; TELEFAX: 202-371-8856  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 655 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; US-08-148-910-12

Query Match 26.4%; Score 832.5; DB 1; Length 655;  
Best Local Similarity 33.2%; Pred. No. 4e-53;  
Matches 192; Conservative 87; Mismatches 216; Indels 83; Gaps 21;



QY 31 ESLDPDWTPOQDYSDYEDYNQEBTSTLTHAENPDW-----YYTEDQA----- 74  
DB 99 QALTEGDRPCRFPRYGRMLHACTSEGSAHRK---WCATTHNYDRDRAGWCVBEATPPP 155  
QY 75 -----DPCQNPCEHGGDC--LVHGSTFTCSCLAPFGSKCKQKQVNTCKDNPCGSGOCL 126  
DB 156 GGPAALDPCASGCLNGGSCSNTQDPOSHYCHSCCPRAFTG-----KD--CGTEKCF 203  
QY 127 ITQSPPPY-----RCVKCHPYTGSCSQV--PVCRPNPCQNGATCSRHK 169  
DB 204 DTRYEYLEGGDRWARVRQGHVEQCEC---FGRTWCEGTRHTACLSSPCLNGTCHLIV 260  
QY 170 RRSKFTACDDQPKGKFCETGSD--CYVGDGYSYRGKQNRVNOHACLWNSHLLLOENY 228  
DB 261 ATGTTVCACPPGFAGRLCNTEPDERCFLNGTGYRGVASTSASGLSCLWNSDLYQELH 320  
QY 229 NMFMEDAETHGIGHNFCRFPDADEKPCWCFIKVTNDKVWEYCDVSACSA--QDVAYPEE 286  
DB 321 VDSVGAALGLGPHAYCRPNDRPWCYV--VKDSALSWEYCRLEACESLTVQLSPDL 379  
QY 287 SPTSPSTKLPGFDCSGCTEIAERKIK--RIYGGFKSTAGKHPQWQASLQSSLPITISMPQGH 345  
DB 380 LATLPEPASPGRQACGRHRKRTFLRPRIIGGSSSLPGSHFPLAAIYIG-----DS 430  
QY 346 FCGGALIHPCWLTAAHCTDIKTRH--LKVVLGDQDLKKEEFHEQSFVEKIFKYSHYNE 403  
DB 431 FCAGSLVHTCWVVSAAHCFSSHPRDSVSVVLGQHFFNRTDVTQTFGIEKYPTILYSV 490  
QY 404 RDEIPHNDIALKLKPVGDGHCALESKYVTKVCLPD--GSPFSGSECHISGNG-VTETGKG 460  
DB 491 FNPSDH-DLVILRLKXKGDRCATRSQFVQPCLPPEPGSTFPAGHKCQIAGWGHLDENYSG 549  
QY 461 -SRQLLDKVKLIANTLNSRQLYDHMDSDSMICAGNLQKPGQDTCCGDSGGPLTCEKDG 519  
DB 550 YSSSLREALVPLVADHKCSSPEVYGADLSPNNLCAAGFYDCK-SDACQDGGGFLACEKNG 608  
QY 520 TYVYGVISWGLECGK--RPGVYTVTKFLNWKATIK 555  
DB 609 VAYLYGIISWGGCGELHKKFPGYVTVRVANYVDWINDRIR 646

US-08-448-937A-12

Query Match 26.4%; Score 832.5; DB 1; Length 655;  
Best Local Similarity 33.2%; Pred. No. 4e-53;  
Matches 192; Conservative 87; Mismatches 216; Indels 83; Gaps 21;

QY 31 ESLDPDWTPOQDYSDYEDYNQEBTSTLTHAENPDW-----YYTEDQA----- 74  
DB 99 QALTEGDRPCRFPRYGRMLHACTSEGSAHRK---WCATTHNYDRDRAGWCVBEATPPP 155  
QY 75 -----DPCQNPCEHGGDC--LVHGSTFTCSCLAPFGSKCKQKQVNTCKDNPCGSGOCL 126  
DB 156 GGPAALDPCASGCLNGGSCSNTQDPOSHYCHSCCPRAFTG-----KD--CGTEKCF 203  
QY 127 ITQSPPPY-----RCVKCHPYTGSCSQV--PVCRPNPCQNGATCSRHK 169  
DB 204 DTRYEYLEGGDRWARVRQGHVEQCEC---FGRTWCEGTRHTACLSSPCLNGTCHLIV 260  
QY 170 RRSKFTACDDQPKGKFCETGSD--CYVGDGYSYRGKQNRVNOHACLWNSHLLLOENY 228  
DB 261 ATGTTVCACPPGFAGRLCNTEPDERCFLNGTGYRGVASTSASGLSCLWNSDLYQELH 320  
QY 229 NMFMEDAETHGIGHNFCRFPDADEKPCWCFIKVTNDKVWEYCDVSACSA--QDVAYPEE 286  
DB 321 VDSVGAALGLGPHAYCRPNDRPWCYV--VKDSALSWEYCRLEACESLTVQLSPDL 379  
QY 287 SPTSPSTKLPGFDCSGCTEIAERKIK--RIYGGFKSTAGKHPQWQASLQSSLPITISMPQGH 345  
DB 380 LATLPEPASPGRQACGRHRKRTFLRPRIIGGSSSLPGSHFPLAAIYIG-----DS 430  
QY 346 FCGGALIHPCWLTAAHCTDIKTRH--LKVVLGDQDLKKEEFHEQSFVEKIFKYSHYNE 403  
DB 431 FCAGSLVHTCWVVSAAHCFSSHPRDSVSVVLGQHFFNRTDVTQTFGIEKYPTILYSV 490  
QY 404 RDEIPHNDIALKLKPVGDGHCALESKYVTKVCLPD--GSPFSGSECHISGNG-VTETGKG 460  
DB 491 FNPSDH-DLVILRLKXKGDRCATRSQFVQPCLPPEPGSTFPAGHKCQIAGWGHLDENYSG 549  
QY 461 -SRQLLDKVKLIANTLNSRQLYDHMDSDSMICAGNLQKPGQDTCCGDSGGPLTCEKDG 519  
DB 550 YSSSLREALVPLVADHKCSSPEVYGADLSPNNLCAAGFYDCK-SDACQDGGGFLACEKNG 608  
QY 520 TYVYGVISWGLECGK--RPGVYTVTKFLNWKATIK 555  
DB 609 VAYLYGIISWGGCGELHKKFPGYVTVRVANYVDWINDRIR 646

RESULT 2

US-08-448-937A-12  
Sequence 12, Application US/08448937A  
Patent No. 5677164  
GENERAL INFORMATION:  
APPLICANT: Takeshi SHIMOMURA et al.  
TITLE OF INVENTION: No. 5677164el Protein and Gene Encoding Said Protein  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch,  
MEDIUM TYPE: 500 Kb Storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/448,937A  
FILING DATE: May 24, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/148,910  
FILING DATE: No. 5677164ember 5, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367

RESULT 3

5244676-5  
Patent No. 5244676  
APPLICANT: BELL, LESLIE D.;MAYER, ERNEST J.;PALMIER, MARK O.  
TOLUNAY, H.ESER;WARREN, THOMAS G.;MUN, TZE-CHIEIN  
TITLE OF INVENTION: MODIFIED TISSUE PLASMINOGEN ACTIVATOR  
WITH MODIFIED GLYCOSYLATION SITE  
NUMBER OF SEQUENCES: 9  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/203,047  
FILING DATE: 06-JUN-1988  
SEQ ID NO:15;

REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX: 202-371-8856  
TELEX:  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 655 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE: human  
ORGANISM: human  
US-08-448-937A-12

QY 31 ESLDPDWTPOQDYSDYEDYNQEBTSTLTHAENPDW-----YYTEDQA----- 74  
DB 99 QALTEGDRPCRFPRYGRMLHACTSEGSAHRK---WCATTHNYDRDRAGWCVBEATPPP 155  
QY 75 -----DPCQNPCEHGGDC--LVHGSTFTCSCLAPFGSKCKQKQVNTCKDNPCGSGOCL 126  
DB 156 GGPAALDPCASGCLNGGSCSNTQDPOSHYCHSCCPRAFTG-----KD--CGTEKCF 203  
QY 127 ITQSPPPY-----RCVKCHPYTGSCSQV--PVCRPNPCQNGATCSRHK 169  
DB 204 DTRYEYLEGGDRWARVRQGHVEQCEC---FGRTWCEGTRHTACLSSPCLNGTCHLIV 260  
QY 170 RRSKFTACDDQPKGKFCETGSD--CYVGDGYSYRGKQNRVNOHACLWNSHLLLOENY 228  
DB 261 ATGTTVCACPPGFAGRLCNTEPDERCFLNGTGYRGVASTSASGLSCLWNSDLYQELH 320  
QY 229 NMFMEDAETHGIGHNFCRFPDADEKPCWCFIKVTNDKVWEYCDVSACSA--QDVAYPEE 286  
DB 321 VDSVGAALGLGPHAYCRPNDRPWCYV--VKDSALSWEYCRLEACESLTVQLSPDL 379  
QY 287 SPTSPSTKLPGFDCSGCTEIAERKIK--RIYGGFKSTAGKHPQWQASLQSSLPITISMPQGH 345  
DB 380 LATLPEPASPGRQACGRHRKRTFLRPRIIGGSSSLPGSHFPLAAIYIG-----DS 430  
QY 346 FCGGALIHPCWLTAAHCTDIKTRH--LKVVLGDQDLKKEEFHEQSFVEKIFKYSHYNE 403  
DB 431 FCAGSLVHTCWVVSAAHCFSSHPRDSVSVVLGQHFFNRTDVTQTFGIEKYPTILYSV 490  
QY 404 RDEIPHNDIALKLKPVGDGHCALESKYVTKVCLPD--GSPFSGSECHISGNG-VTETGKG 460  
DB 491 FNPSDH-DLVILRLKXKGDRCATRSQFVQPCLPPEPGSTFPAGHKCQIAGWGHLDENYSG 549  
QY 461 -SRQLLDKVKLIANTLNSRQLYDHMDSDSMICAGNLQKPGQDTCCGDSGGPLTCEKDG 519  
DB 550 YSSSLREALVPLVADHKCSSPEVYGADLSPNNLCAAGFYDCK-SDACQDGGGFLACEKNG 608  
QY 520 TYVYGVISWGLECGK--RPGVYTVTKFLNWKATIK 555  
DB 609 VAYLYGIISWGGCGELHKKFPGYVTVRVANYVDWINDRIR 646

LENGTH: 562  
5244676-5

Query Match 24.9%; Score 786; DB 6; Length 562;  
Best Local Similarity 33.8%; Pred. No. 8.5e-50;  
Matches 181; Conservative 74; Mismatches 177; Indels 104; Gaps 16;  
QY 112 QNTCKNFCGQCLITQSPYYR-----CVCKHPTGSCSQVVPV--CRNPPC 159  
DB 38 QVICRDE---KTQMIYQHQSWLRPLVRSNREVEYWCN---SGRAQCHSVPRKSCSPRC 91  
QY 160 QNGATCSRHRKRSKFTACPOFGKCEI--GSDDCYVGDGYSYRGKMNRTVNOHACLYW 218  
DB 92 FNGGTQCALYPSDFVCCPGFPAKCEIDGNSDCYFGSGAYRTHSTESGASCLPW 151  
QY 219 NSHLLQENYMFMEDAETHGIGEHNFRCNPDADKPMCFIKVTNDKVKWEYCDVSACSA 278  
DB 152 NSMILIGKYTAQNPSAALGLGKHNYCRNPDGAKPMCHV-LKNRRLTWECYCDVPCSE 210  
QY 279 QD-----VAYPEESTEPSTKL----- 295  
DB 211 GNSDCYFNGSAYRGTHSLTESGASCLPWNSMILIGKYTAQNPSAALGLGKHNYCRNP 270  
QY 296 -----FGFDCSGKTEIAERKIKIYGGFKSTAGKHPWQASLQ 332  
DB 271 DGDAPKPMCHVKNRRLTWECYCDVPCSTCGLRQYSQPOF-RKGGLPADIASHPWQAIF 329  
QY 333 SSLPLTISMPQCHFGGALIHPCWVLTAAHCTD--IKTRHLKVLGDODLKKKEEFHEQSF 390  
DB 330 AK--HRRSPGFERFLCGGILSSCWILSAHCFQERPPHLLTVILGRYRVVPGEEQKF 387  
QY 391 RVEKFKYSHYNERDEIPHNDIALKLPVDGHCALSKYKVTCLP--DGSFPGSGECH 448  
DB 388 EVEKIVHKEFD--DDTYNDIALQLKSDSSRCQESSVVRTVCLPAD:QLPDWTECE 445  
QY 449 ISGWGVTETGKG--SRQLLDKVKLIANTLNSRQLYDHMIDDSMICAGNLQKPG----- 501  
DB 446 LSGYKHEALSPFYGERLKEAHRVLYPSRCSCTSHLLNRVTNMLCAGDTRSGGPOANL 505  
QY 502 QDTCQDGGGGLTCEKDGTYVYVGVSWGLEGKGR--PGVTVQVTKFLNWKATIK 555  
DB 506 HDACQDGGGGLVCLNDGEMTLVGIISWGLGCGQKDVGVTKVTNYLDWIRDNR 561

## RESULT 4

Patent No. 5520913  
; APPLICANT: ANDERSON, STEPHEN; BENNETT, WILLIAM F.; BOTSSTEIN.  
; DAVID; HIGGINS, DEBORAH L.; PAONI, NICHOLAS F.; ZOLLER, MARK J.  
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR HAVING  
; ZYMOGENIC PROPERTIES  
; NUMBER OF SEQUENCES: 35  
; APPLICATION NUMBER: US/08/89,451  
; FILING DATE: 06-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 770,510  
; FILING DATE: 03-OCT-1991  
; APPLICATION NUMBER: 384,608  
; FILING DATE: 24-JUL-1989  
; APPLICATION NUMBER: 240,856  
; FILING DATE: 02-SEP-1988  
; SEQ ID NO: 1  
; LENGTH: 527  
5520913-1

Query Match 24.7%; Score 777.5; DB 6; Length 527;  
Best Local Similarity 35.2%; Pred. No. 3.3e-49;  
Matches 183; Conservative 76; Mismatches 176; Indels 85; Gaps 20;  
QY 77 CQNPCEHGGDC--LVHGTFTTCSCLAPSGNKKQ-KVQNTCKNPP--CGRG----- 123  
DB 51 CSEPRCFNGGTCCQALYFDFVCQCPGEPGAGKCCSIDTRATCYEDQGISYRGTWSTAESG 110

## RESULT 5

US-08-811-949-39  
; Sequence 39, Application US/08811949  
; Patent No. 5840533  
; GENERAL INFORMATION:  
; APPLICANT: NIWA, MINBO  
; APPLICANT: SAITO, YOSHIMASA  
; APPLICANT: SASAKI, HITOSHI  
; APPLICANT: HAYASHI, MASAKO  
; APPLICANT: NOTANI, JOUJI  
; APPLICANT: KOBAYASHI, MASAKAZU  
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR  
; NUMBER OF SEQUENCES: 67  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,949  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 18-966-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:

QY 124 -QCLITQSPYYRCVCKHPTGTPSCSQVVPVCRPNPQ-----NGATCSRHRKRSKFTAC 178  
DB 111 ABCTNWS-----SALAKPYSGR-----RPDAIRLGLGNHNYCRNPDSDSKPWCYV 157  
QY 179 PQQFK-GKP-----CEIGSDDCYVGDGYSYRGKMNRTVNOHACLYWNSHLLQENY 228  
DB 158 ---FRAGKYSSEFCSTPACSEGNSCYFGNGSAYRGTHSLTESGASCLPWNSMILRHQKY 214  
QY 229 NMFMEDAETHGIGEHNFRCNPDADKPMCFIKVTNDKVKWEYCDVSACSAQDVAYPESP 288  
DB 215 TAQNSAQAALGLGKHNYCRNPDGAKPMCHV-LKNRRLTWECYCDVPCSCS----- 262  
QY 289 TEPSTKLPDFDCSGKTEIAERKIKIYGGFKSTAGKHPWQASLQSSLPITISMPQHFQ 348  
DB 263 -----TCGLRQYSQPOF-RKGGLPADIASHPWQAIFAK--HRRSPGGERHLCG 308  
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVLGDODLKKKEEFHEQSFVEKIFKYSHYNERDE 406  
DB 309 GILISSCWILSAHCFQERPPHLLTVILGRYRVVPGEEQKFEVEKIVHKEFD--DD 366  
QY 407 IPHNDIALKLPVDGHCALSKYKVTCLP--DGSFPGSGECHISGWGVTETGKG--SR 462  
DB 367 TYDNDIALQLKSDSSRCQESSVVRTVCLPAD:QLPDWTECELSGYKHEALSPFYSE 426  
QY 463 QLDDAKVKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----QDTCQDGGGGLTCEK 517  
DB 427 RUKEAHRVLYPSRCSCTSHLLNRVTNMLCAGDTRSGGPOANLHDACQDGGGGLVCLN 486  
QY 518 DGTYYVGVSWGLEGKGR--PGVTVQVTKFLNWKATIK 555  
DB 487 DGRMTLVGIISWGLGCGQKDVGVTKVTNYLDWIRDNR 526

```
;
; LENGTH: 527 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; MOLECULE TYPE: protein
US-08-811-949-39

Query Match      24.3%   Score 767.5; DB 2; Length 527;
Best Local Similarity 35.08; Pred. No. 1.8e-48; Indels 85; Gaps 20;
Matches 182; Conservative 77; Mismatches 176;

QY 77 CQNPCEHGGDC--LVHGSTFTSCCLAPFSGNKCQ-KVQNTCKDNP--CGRG-----123
Db 51 CSEPRCNGGTCCQALYFSDVFCQCEGAGKCEIDTRATCYEDQGISYRGTWSTAESG 110
QY 124 -QCLITQSPYYRCVCKHPTGSCSQVVPVCPNQCQ----NGATCSRHKRSKFTCAC 178
Db 111 AECTNWN-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDROSKPWCYV 157
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYRGKMNRTVNHQACLYWNSHLLQENY 228
Db 158 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKYV 214
QY 229 NMFMEDAEHIGEHNECRNPDADEKPCWCFIKVTNDKVWEYCDVSAQDVAYPEESP 288
Db 215 TAQNPQAALGLGNHNYCRNPDGDAKPCWCHV-LKNERLTWEYCDVPSCS-----262
QY 289 TEPSTKLPGFDSCKTIAERKIKRIYGGFKSTAGKHPQASQSSLPITISMPQGHFCG 348
Db 263 -----TCGLRQYSQPF-RIGGLFADIASHPWQAALPAK--HRRSPGERFLCG 308
QY 349 GALIHPQWLVTAHCTD--IKTRHLKVVLGDQDLKKEEFHEQSPRVEKIFKYSHYNERDE 406
Db 309 GILISSCWILSAAHCFQERPPPHLTVILGRTYRVVPGEEQFEVEKIVHKEFD--DD 366
QY 407 IPHNDIALLLKPVGDHCALESKYKTVCLP--DGSFPSSGECHISGWGVTETGK--SR 462
Db 367 TYNDIALLLQKSDSSRCAQESSVTVCLPPADLQLPDWTCELSGYGHEALSFPYSE 426
QY 463 QLLDAKVKLIANTLNSRQLYDHMDSDMTCAGNLQKPG-----QDTCQDSSGGPLTCEK 517
Db 427 RLKEARVRLYPSRSTSHLLNRTVTDNMLCAGDTRSGGPQANLHDACQDSSGGPLVCLN 486
QY 518 DGTYYVYGVISWGLECKGR--PGVYTVTKFLNWKATIK 555
Db 487 DGRMTLVGLISWGLGCGQKDVPGVYTKVTNYLDWIRDNR 526

RESULT 6
US-07-609-510B-16
; Sequence 16, Application US/07609510B
; Patent No. 5326700
; GENERAL INFORMATION:
; APPLICANT: Berg et al.
; TITLE OF INVENTION: Method for Altering Post-Translational Processing of Tissue P
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: IN.
; COUNTRY: U.S.A.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,510B
; FILING DATE: 19901106
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
```

```
;
; LENGTH: 527 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-609-510B-16

Query Match      24.3%   Score 766.5; DB 1; Length 527;
Best Local Similarity 35.08; Pred. No. 2.1e-48; Indels 85; Gaps 20;
Matches 182; Conservative 77; Mismatches 176;

QY 77 CQNPCEHGGDC--LVHGSTFTSCCLAPFSGNKCQ-KVQNTCKDNP--CGRG-----123
Db 51 CSEPRCNGGTCCQALYFSDVFCQCEGAGKCEIDTRATCYEDQGISYRGTWSTAESG 110
QY 124 -QCLITQSPYYRCVCKHPTGSCSQVVPVCPNQCQ----NGATCSRHKRSKFTCAC 178
Db 111 AECTNWN-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDROSKPWCYV 157
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYRGKMNRTVNHQACLYWNSHLLQENY 228
Db 158 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKYV 214
QY 229 NMFMEDAEHIGEHNECRNPDADEKPCWCFIKVTNDKVWEYCDVSAQDVAYPEESP 288
Db 215 TAQNPQAALGLGNHNYCRNPDGDAKPCWCHV-LKNERLTWEYCDVPSCS-----262
QY 289 TEPSTKLPGFDSCKTIAERKIKRIYGGFKSTAGKHPQASQSSLPITISMPQGHFCG 348
Db 263 -----TCGLRQYSQPF-RIGGLFADIASHPWQAALPAK--HRRSPGERFLCG 308
QY 349 GALIHPQWLVTAHCTD--IKTRHLKVVLGDQDLKKEEFHEQSPRVEKIFKYSHYNERDE 406
Db 309 GILISSCWILSAAHCFQERPPPHLTVILGRTYRVVPGEEQFEVEKIVHKEFD--DD 366
QY 407 IPHNDIALLLKPVGDHCALESKYKTVCLP--DGSFPSSGECHISGWGVTETGK--SR 462
Db 367 TYNDIALLLQKSDSSRCAQESSVTVCLPPADLQLPDWTCELSGYGHEALSFPYSE 426
QY 463 QLLDAKVKLIANTLNSRQLYDHMDSDMTCAGNLQKPG-----QDTCQDSSGGPLTCEK 517
Db 427 RLKEARVRLYPSRSTSHLLNRTVTDNMLCAGDTRSGGPQANLHDACQDSSGGPLVCLN 486
QY 518 DGTYYVYGVISWGLECKGR--PGVYTVTKFLNWKATIK 555
Db 487 DGRMTLVGLISWGLGCGQKDVPGVYTKVTNYLDWIRDNR 526

RESULT 7
PCT-US91-01025A-2
; Sequence 2, Application PC/TUS9101025A
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; TITLE OF INVENTION: Tissue Plasminogen Activator Having Fibrin
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/01025A
; FILING DATE: 19910214
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: 07/486,657
; FILING DATE: 1 March 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 454P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 527 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; PCT-US91-01025A-2

Query Match          24.3%; Score 766.5; DB 5; Length 527;
Best Local Similarity 35.0%; Pred. No. 2.1e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ-KVQNTCKDNP--CGRG-----123
Db 51 CSEPRCFNGGTCQQALYFSDVFCQCPGFAKCEIDTRATCYEDQGISYRGTWSTAESG 110
QY 124 -QCLITOSPYYRCVCKHPYTGPSCSQVVPVCRPNPCQ----NGATCSRHKRSKFTCAC 178
Db 111 AECTNWS-----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPD RDSKPCYV 157
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMNRTVNOHACLWNHLLQENY 228
Db 158 ---FRAGKYSSEFCSTPACSEGSDCYFGNGSAYRGTHSLTESGASCLPWNMILGKYV 214
QY 229 NMFMEDAETHGIGEHNFERNPDADKPCWFIKVTNDKVWEYCDVSACSAQDVAYPEESP 288
Db 215 TAQNPQAALGLGKHNYCRNPDGAKPWCHV-LKNRRLTWECYCDVPSCS-----262
QY 289 TEPSTKLPDGSCKTIAERKIRIYGGFKSTAGKHPWQASLOSPLTISMPQHFCG 348
Db 263 -----TCGLRYSQPF-RIKGGLFADIASHPWQAIFAK--HRRSEGERFLCG 308
QY 349 GALIHPCWVLTAHCTD--IKTRHLKVVLGDQDLKKEEFHQSPRVEKIPKYSHYNERDE 406
Db 309 GILLISSCWILSAACHFCQERFPFPHLTIVILGRTYRVVPGEEQKFEVEKIVHKEFD--DD 366
QY 407 IPHNDIALKLKVPDGHCALESKYVTVCLP--DGSFPGSGECHISGNGVTETGK--SR 462
Db 367 TYDNDIALQLKSDSSRCAQESSVVRTVCLFPADQLPDMTECELSGYGKHEALSPPFYS 426
QY 463 QLLDAKVLIANTLCNSRQLYDHMDISMICAGNLQKPG-----QDTCQDGGSGPLTCEK 517
Db 427 RLKEAHVLYPSSRCTSHLLNRTVTDNMLCAGTRSGGPOANLHDACQDGGSGPLVCLN 486
QY 518 DGTYYVYGVISWGLECGKR--PGVYTVTKFLNMIKATIK 555
Db 487 DGRMTLVGIISWGLGCGQKQDVGYYTKVTNYLDWIRDNR 526
```

```
RESULT 8
5185259-8
; Patent No. 5185259
; APPLICANT: GOEDDEL, DAVID V.; KOHR, WILLIAM J.; PENNICA, DIANE;
; VEHAR, GORDON A.
; TITLE OF INVENTION: TRUNCATED HUMAN TISSUE PLASMINOGEN
; ACTIVATOR
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/489,855
; FILING DATE: 02-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 12,694
; FILING DATE: 09-FEB-1987
; APPLICATION NUMBER: 483,052
```

```
; FILING DATE: 07-APR-1983
; APPLICATION NUMBER: 398,003
; FILING DATE: 14-JUL-1982
; APPLICATION NUMBER: 374,860
; FILING DATE: 05-MAY-1982
; SEQ ID NO:8:
; LENGTH: 527
5185259-8

Query Match          24.3%; Score 766.5; DB 6; Length 527;
Best Local Similarity 35.0%; Pred. No. 2.1e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ-KVQNTCKDNP--CGRG-----123
Db 51 CSEPRCFNGGTCQQALYFSDVFCQCPGFAKCEIDTRATCYEDQGISYRGTWSTAESG 110
QY 124 -QCLITOSPYYRCVCKHPYTGPSCSQVVPVCRPNPCQ----NGATCSRHKRSKFTCAC 178
Db 111 AECTNWS-----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPD RDSKPCYV 157
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMNRTVNOHACLWNHLLQENY 228
Db 158 ---FRAGKYSSEFCSTPACSEGSDCYFGNGSAYRGTHSLTESGASCLPWNMILGKYV 214
QY 229 NMFMEDAETHGIGEHNFERNPDADKPCWFIKVTNDKVWEYCDVSACSAQDVAYPEESP 288
Db 215 TAQNPQAALGLGKHNYCRNPDGAKPWCHV-LKNRRLTWECYCDVPSCS-----262
QY 289 TEPSTKLPDGSCKTIAERKIRIYGGFKSTAGKHPWQASLOSPLTISMPQHFCG 348
Db 263 -----TCGLRYSQPF-RIKGGLFADIASHPWQAIFAK--HRRSEGERFLCG 308
QY 349 GALIHPCWVLTAHCTD--IKTRHLKVVLGDQDLKKEEFHQSPRVEKIPKYSHYNERDE 406
Db 309 GILLISSCWILSAACHFCQERFPFPHLTIVILGRTYRVVPGEEQKFEVEKIVHKEFD--DD 366
QY 407 IPHNDIALKLKVPDGHCALESKYVTVCLP--DGSFPGSGECHISGNGVTETGK--SR 462
Db 367 TYDNDIALQLKSDSSRCAQESSVVRTVCLFPADQLPDMTECELSGYGKHEALSPPFYS 426
QY 463 QLLDAKVLIANTLCNSRQLYDHMDISMICAGNLQKPG-----QDTCQDGGSGPLTCEK 517
Db 427 RLKEAHVLYPSSRCTSHLLNRTVTDNMLCAGTRSGGPOANLHDACQDGGSGPLVCLN 486
QY 518 DGTYYVYGVISWGLECGKR--PGVYTVTKFLNMIKATIK 555
Db 487 DGRMTLVGIISWGLGCGQKQDVGYYTKVTNYLDWIRDNR 526
```

```
RESULT 9
US-08-811-949-43
; Sequence 43, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 562 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-949-43

Query Match 24.3%; Score 766.5; DB 2; Length 562;
Best Local Similarity 35.0%; Pred. No. 2.3e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFGNKKQ-KVQNTCKDNP--CGRG-----123
Db 86 CSEPRCENGTCQQAIFYDFVQCPEGFAGKCEIDTRATCYEDQGISYRGTWSTAEG 145
QY 124 -QCLITQSPYYRCVKHPYGPSCSQVVPVCRNPQO---NGATCSRHKRSKFTCAC 178
Db 146 AECTNWS---SALAQKPYSGR-----RPDAIRLGLGNHNYCRNDRSKPWCYV 192
QY 179 PDQFK-GKF-----CEIGSDCYVGDGYSGKGNRTVNOHACLYWNSHLLQENY 228
Db 193 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAVRGTHSLTESGASCLPWNMILIGKY 249
QY 229 NMFEDAEATHGIGEHNFRCNPDABEKPCFKIVTNDKVKWEYCDVSACSADQVAYPESP 288
Db 250 TAQNPQAALGLGKHNYCRNPDGDAKPWCHV-LKNRLTWECYDVPSCS-----297
QY 289 TEPSTKLPGFDSGCKTETABRKIKIYGGFKSTAGKHPQASLOSLLPLTISMPQGHFCG 348
Db 298 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIAFAK--HRRSPGERFLCG 343
QY 349 GALIHPCWLTAAHCTD--IKTRHLKVLVGDQDLKKEEFHQSFVEKIFKYSHYNERDE 406
Db 344 GILISSCWILSAAHCFQERFPFHLLTVILGRYRVVPGEEQKFEVEKIYVHKEFD--DD 401
QY 407 IPHNDIALKLKPVDPVGHCALESKYVKTCLP--DGSFPGSGECHISGWGTETGKG--SR 462
Db 402 TYDNDIALQLKSDSRCAQESSVVRTVCLPPADLQLPDWTCELSGVGKHEALSPFYSE 461
QY 463 QLLDAVKKLIANTLNCNSRLYDHMIDDSMICAGNLQKPG-----QDTCCGDSGGPLTCEK 517
Db 462 RLKEAHRVLYPSSRCTSOHLLNRTVTDNMLCAGDTRSGGPOANLHDACQSDSGGLVCLN 521
QY 518 DGYVYVGVISWGLECGKR--PGVYQVTKFLNWKATIK 555
Db 522 DGRMTVLGLISWGLECGQKDPGVYTKVTNVLWDNRDMR 561

```

```

RESULT 10
US-08-560-098A-50
; Sequence 50, Application US/08560098A
; Patent No. 5976841
; GENERAL INFORMATION:
; APPLICANT: WENDEL, Stephan
; APPLICANT: HEINZEL-WIELAND, Regina
; APPLICANT: STEFFENS, Gerd Josef
; TITLE OF INVENTION: Proteins having Fibrinolytic and
; TITLE OF INVENTION: Coagulation-inhibiting Properties

```

```

; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Evenson, McKeown, Edwards & Lenahan
; STREET: 1200 G Street, N.W., Suite 700
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560,098A
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P 44 40 892.7
; FILING DATE: 17-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: EVANS, Joseph D.
; REGISTRATION NUMBER: 26,269
; REFERENCE/DOCKET NUMBER: 148/42448
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 562 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-560-098A-50

```

```

Query Match 24.3%; Score 766.5; DB 2; Length 562;
Best Local Similarity 35.0%; Pred. No. 2.3e-48;
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFGNKKQ-KVQNTCKDNP--CGRG-----123
Db 86 CSEPRCENGTCQQAIFYDFVQCPEGFAGKCEIDTRATCYEDQGISYRGTWSTAEG 145
QY 124 -QCLITQSPYYRCVKHPYGPSCSQVVPVCRNPQO---NGATCSRHKRSKFTCAC 178
Db 146 AECTNWS---SALAQKPYSGR-----RPDAIRLGLGNHNYCRNDRSKPWCYV 192
QY 179 PDQFK-GKF-----CEIGSDCYVGDGYSGKGNRTVNOHACLYWNSHLLQENY 228
Db 193 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAVRGTHSLTESGASCLPWNMILIGKY 249
QY 229 NMFEDAEATHGIGEHNFRCNPDABEKPCFKIVTNDKVKWEYCDVSACSADQVAYPESP 288
Db 250 TAQNPQAALGLGKHNYCRNPDGDAKPWCHV-LKNRLTWECYDVPSCS-----297
QY 289 TEPSTKLPGFDSGCKTETABRKIKIYGGFKSTAGKHPQASLOSLLPLTISMPQGHFCG 348
Db 298 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIAFAK--HRRSPGERFLCG 343
QY 349 GALIHPCWLTAAHCTD--IKTRHLKVLVGDQDLKKEEFHQSFVEKIFKYSHYNERDE 406
Db 344 GILISSCWILSAAHCFQERFPFHLLTVILGRYRVVPGEEQKFEVEKIYVHKEFD--DD 401
QY 407 IPHNDIALKLKPVDPVGHCALESKYVKTCLP--DGSFPGSGECHISGWGTETGKG--SR 462
Db 402 TYDNDIALQLKSDSRCAQESSVVRTVCLPPADLQLPDWTCELSGVGKHEALSPFYSE 461
QY 463 QLLDAVKKLIANTLNCNSRLYDHMIDDSMICAGNLQKPG-----QDTCCGDSGGPLTCEK 517
Db 462 RLKEAHRVLYPSSRCTSOHLLNRTVTDNMLCAGDTRSGGPOANLHDACQSDSGGLVCLN 521
QY 518 DGYVYVGVISWGLECGKR--PGVYQVTKFLNWKATIK 555

```



RESULT 13  
5185259-3  
; Patent No. 5185259  
; APPLICANT: GOEDDEL, DAVID V.; KOHR, WILLIAM J.; PENNICA, DIANE;  
; VEHAR, GORDON A.  
; TITLE OF INVENTION: TRUNCATED HUMAN TISSUE PLASMINOGEN  
; ACTIVATOR  
; NUMBER OF SEQUENCES: 15  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/489,855  
; FILING DATE: 02-MAR-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 12,694  
; FILING DATE: 09-FEB-1987  
; APPLICATION NUMBER: 483,052  
; FILING DATE: 07-APR-1983  
; APPLICATION NUMBER: 398,003  
; FILING DATE: 14-JUL-1982  
; APPLICATION NUMBER: 374,860  
; FILING DATE: 05-MAY-1982  
; SEQ ID NO:3:  
; LENGTH: 562  
5185259-3  
Query Match 24.3%; Score 766.5; DB 6; Length 562;  
Best Local Similarity 35.0%; Pred. No. 2.3e-48;  
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;  
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ--KVQNTCKDNP--CGRG-----123  
Db 86 CSEPRCFNGGTCCQALYFSDVFCQPEGFAGKCEIDTRATCYEDQGISVGTWSTAESG 145  
QY 124 -QCLITQSPPYRYRCVCKHPYTGPCSQVVPVCRNPQ-----NGATCSRHKRSKFTCAC 178  
Db 146 AECTNWS-----SALAQKPYSGR-----RDAIRLGLGNHNYCRNPDSDKPCYV 192  
QY 179 PDQFK-GKF-----CEIGSDDCYVGDYSGYRGKXNRVNOHACLYWNSHLLQENY 228  
Db 193 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNSMILIGKVY 249  
QY 229 NFMEDAEATHGIGHNFCRNPDADEKPCWCFIKVNDKVWYCDVSAQSDVAYPEESP 288  
Db 250 TAQNPQAALGLGKXNYCRNPDGAKPWCHV-LXNRRLTWYCDVPCS-----297  
QY 289 TEPSTKLPFCPSGCKTEIAERKIRIYGGFKSTAGKHPQWASLOSSLPITSMPOGHFCG 348  
Db 298 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIFAK--HRRSPGERFLCG 343  
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVVLGQDLKKEEFHQSPRVEKIFKYSHYNERDE 406  
Db 344 GILISSCWILSAHCFQERFPFPHLTVILGRTYRVVPGEEBQKEVEKIVHKEFD--DD 401  
QY 407 IPHNDIALKLPVDGHCALSKYKVTCLP--DGSPPSGSECHISGMVGTETGK--SR 462  
Db 402 TYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGKEALSPPYSE 461  
QY 463 QLLDAKVKLIANTLNSRQLYDHMDSDMI CAGNLQKPG-----QDTCCGDSGGPLTCEK 517  
Db 462 RLKEARVLYPSRSTCSHLLNRTVTDNMLCAGDTRSGGPOANLHDACQDSGGPLVCLN 521  
QY 518 DGTYYVYGVISVGLGCGKR--PGVYTVQTKFLNMIKATIK 555  
Db 522 DGRMTLVGIISVGLGCGQKDVGVYTKVTNYLDWIRDNR 561  
RESULT 14  
5200340-2  
; Patent No. 5200340  
; APPLICANT: FOSTER, DONALD C.; MULVATHILL, EILEEN R.; O'HARA,  
; PATRICK J.; PINGEL, KURT; YOSHITAKE, SHINJI  
; TITLE OF INVENTION: THROMBIN-ACTIVATED TISSUE PLASMINOGEN  
; ACTIVATORS  
; NUMBER OF SEQUENCES: 34  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/782,686  
; FILING DATE: 01-OCT-1985  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 656,770  
; FILING DATE: 01-OCT-1984  
; SEQ ID NO:2:  
; LENGTH: 562  
5200340-2  
Query Match 24.3%; Score 766.5; DB 6; Length 562;  
Best Local Similarity 35.0%; Pred. No. 2.3e-48;  
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;  
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ--KVQNTCKDNP--CGRG-----123

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/53,412  
; FILING DATE: 22-MAY-1987  
; SEQ ID NO:2:  
; LENGTH: 562  
5200340-2  
Query Match 24.3%; Score 766.5; DB 6; Length 562;  
Best Local Similarity 35.0%; Pred. No. 2.3e-48;  
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;  
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ--KVQNTCKDNP--CGRG-----123  
Db 86 CSEPRCFNGGTCCQALYFSDVFCQPEGFAGKCEIDTRATCYEDQGISVGTWSTAESG 145  
QY 124 -QCLITQSPPYRYRCVCKHPYTGPCSQVVPVCRNPQ-----NGATCSRHKRSKFTCAC 178  
Db 146 AECTNWS-----SALAQKPYSGR-----RDAIRLGLGNHNYCRNPDSDKPCYV 192  
QY 179 PDQFK-GKF-----CEIGSDDCYVGDYSGYRGKXNRVNOHACLYWNSHLLQENY 228  
Db 193 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNSMILIGKVY 249  
QY 229 NFMEDAEATHGIGHNFCRNPDADEKPCWCFIKVNDKVWYCDVSAQSDVAYPEESP 288  
Db 250 TAQNPQAALGLGKXNYCRNPDGAKPWCHV-LXNRRLTWYCDVPCS-----297  
QY 289 TEPSTKLPFCPSGCKTEIAERKIRIYGGFKSTAGKHPQWASLOSSLPITSMPOGHFCG 348  
Db 298 -----TCGLRQYSQPF-RIKGGLFADIASHPWQAIFAK--HRRSPGERFLCG 343  
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVVLGQDLKKEEFHQSPRVEKIFKYSHYNERDE 406  
Db 344 GILISSCWILSAHCFQERFPFPHLTVILGRTYRVVPGEEBQKEVEKIVHKEFD--DD 401  
QY 407 IPHNDIALKLPVDGHCALSKYKVTCLP--DGSPPSGSECHISGMVGTETGK--SR 462  
Db 402 TYDNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGKEALSPPYSE 461  
QY 463 QLLDAKVKLIANTLNSRQLYDHMDSDMI CAGNLQKPG-----QDTCCGDSGGPLTCEK 517  
Db 462 RLKEARVLYPSRSTCSHLLNRTVTDNMLCAGDTRSGGPOANLHDACQDSGGPLVCLN 521  
QY 518 DGTYYVYGVISVGLGCGKR--PGVYTVQTKFLNMIKATIK 555  
Db 522 DGRMTLVGIISVGLGCGQKDVGVYTKVTNYLDWIRDNR 561  
RESULT 15  
5344773-2  
; Patent No. 5344773  
; APPLICANT: WEI, CHA-MER; HSIUNG, NANCY; REDDY, VERMURI B.;  
; LEMONT, JEFFREY F.; DACKOWSKI, WILLIAM; DOUGLAS, RICHARD;  
; COLE, EDWARD S.; PURCELL JR., RICHARD D.; LAU, DAVID TAI-YUI  
; TITLE OF INVENTION: HUMAN UTERINE TISSUE PLASMINOGEN  
; ACTIVATOR PRODUCED BY RECOMBIANT DNA  
; NUMBER OF SEQUENCES: 6  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/782,686  
; FILING DATE: 01-OCT-1985  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 656,770  
; FILING DATE: 01-OCT-1984  
; SEQ ID NO:2:  
; LENGTH: 562  
5344773-2  
Query Match 24.3%; Score 766.5; DB 6; Length 562;  
Best Local Similarity 35.0%; Pred. No. 2.3e-48;  
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;  
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKQ--KVQNTCKDNP--CGRG-----123



Db 86 CSEPRCNGGTCQALYFDFVQCPEGFAGKCEIDTRATCYEDQGISYRGTWSTAEG 145  
QY 124 -OCLITQSPYRCVCKHPYGPSCSVQVPCRENPCQ---NGATCSRHKRSKFTCAC 178  
Db 146 AECTNWN---SALAOKPYSGR-----RDAIRLGLGNHNYCNDRDSKPCYV 192  
QY 179 PDQFK-GXF-----CEIGSDCYVGDGYRGKXNRTVNOHACLYWNSHLLQENTY 228  
Db 193 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAYRGTHSLTESGASCLPWNSMILIGKY 249  
QY 229 NMFEDAETHGIGEHNFCEPDDEKWCFFIKVTNDKVKWEYCDVSACSADQVAYPEEP 288  
Db 250 TAQNPSAQLGLGKHNYCRNPDGAKPWCHV-LKNRRLTWIYCDVPSCS----- 297  
QY 289 TEPSTKLPFGDFSCGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSLEPLTISMPOGHFCG 348  
Db 298 -----TCGLRQYSQPF-RKGGLPADIASHPWQAAIPAK-HRRSPGERFLCG 343  
QY 349 GALIHPCWLTAAHCTD--IKTRHLKVVLGQDLKXBEFHEQSPRVEKIFKYSHYNERDE 406  
Db 344 GILISSCWILSAAHCFQERFPFPHLTVILGRTRYVVPGEEOQFEVEKYIVHKEFD--DD 401  
QY 407 IPHNDIALLLKLPVDGHCALESKYVKTCLP--DGSPPSGSECHISGWGTETGKG--SR 462  
Db 402 TYDNDIALLLKLSDSRCAQSSVVRTVCLPADLQDPDTECELSGYGKHEALSIFYSE 461  
QY 463 QLLDAKYXLANTLNSRQLYDHMIDDMSICAGNLQKEG-----OPTCGDSGGPLTCEK 517  
Db 462 RLKEAHVRLYPSSRCTSQHLLNRTVTDNMLCAGDTRSGGPOANLHDACQDSGGPLVCLN 521  
QY 518 DGTIVYVYGIVSWGLECGKR--PGVYTVTKFLNMIKATIK 555  
Db 522 DGRNTLVGLIISWGLCGGQKDVGVYTKVTNYLDWIRDNR 561

Search completed: May 24, 2004, 09:50:30  
Job time : 19.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 09:49:06 ; Search time 42 Seconds  
(without alignments)  
3718.956 Million cell updates/sec

Title: US-09-912-559-3

Perfect score: 3154

Sequence: 1 MFARMSDLHVLLMALVYKLT.....TQVTKFLNWKATIKSESQF 560

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3154	100.0	560	9	US-09-912-559-3
2	3154	100.0	560	14	US-10-172-712-32
3	3154	100.0	560	15	US-10-391-215-5
4	3151	99.9	560	15	US-10-391-215-7
5	3146	99.7	560	15	US-10-391-215-6
6	3143	99.7	560	9	US-09-912-559-4
7	3143	99.7	560	15	US-10-391-215-8
8	832.5	26.4	655	14	US-10-172-712-28
9	767.5	24.3	562	9	US-09-974-298-145
10	766.5	24.3	527	10	US-09-987-457-18
11	766.5	24.3	527	10	US-09-987-455-19
12	766.5	24.3	527	12	US-10-432-842-1
13	766.5	24.3	527	15	US-10-360-101-203
14	766.5	24.3	562	9	US-09-969-271-7
15	766.5	24.3	562	12	US-10-411-037-26

```
16 766.5 24.3 562 12 US-10-411-026-26
17 766.5 24.3 562 14 US-10-193-656-8
18 766.5 24.3 562 14 US-10-443-701-4
19 766.5 24.3 562 16 US-10-410-962-26
20 766.5 24.3 562 16 US-10-411-049-26
21 753.5 23.9 650 15 US-10-401-077-1
22 741 23.5 615 10 US-09-858-909-2
23 741 23.5 615 14 US-10-172-712-30
24 741 23.5 615 16 US-10-449-132-2
25 719 22.8 372 9 US-09-084-491A-3
26 719 22.8 372 13 US-10-103-704-3
27 709.5 22.5 354 10 US-09-987-457-10
28 709.5 22.5 354 10 US-09-987-455-11
29 709.5 22.5 377 10 US-09-987-455-8
30 690 21.6 433 12 US-10-087-192-591
31 662.5 21.0 411 15 US-10-407-821-2
32 661.5 21.0 411 9 US-09-880-503-3
33 661.5 21.0 431 12 US-10-411-037-34
34 661.5 21.0 431 12 US-10-411-026-34
35 661.5 21.0 431 13 US-10-076-421-2
36 661.5 21.0 431 14 US-10-171-311-184
37 661.5 21.0 431 14 US-10-301-822-161
38 661.5 21.0 431 14 US-10-131-985-21
39 661.5 21.0 431 15 US-10-295-027-414
40 661.5 21.0 431 15 US-10-295-027-1275
41 661.5 21.0 431 16 US-10-410-962-34
42 661.5 21.0 431 16 US-10-411-049-34
43 661.5 21.0 437 12 US-10-087-192-594
44 660.5 20.9 431 14 US-10-247-671-149
45 660 20.9 812 12 US-09-825-751A-71
```

#### ALIGNMENTS

#### RESULT 1

```
US-09-912-559-3
; Sequence 3, Application US/09912559
; Patent No. US20020142316A1
; GENERAL INFORMATION:
; APPLICANT: ROEMISCH, JURGEN
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: FEUSNER, ANNETTE
; APPLICANT: LANG WIEGAND
; APPLICANT: WEIMER, THOMAS
; APPLICANT: BECKER, MARGRET
; APPLICANT: NERLICH, CLAUDIA
; APPLICANT: MUTH-NAUMANN, GUDRUN
; TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND
; DETECTION METHODS USING SPECIFIC ANTIBODIES
; FILE REFERENCE: 06478.1457
; CURRENT APPLICATION NUMBER: US/09/912,559
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-912-559-3
```

Query Match 100.0%; Score 3154; DB 9; Length 560;  
Best Local Similarity 100.0%; Pred. No. 9.1e-241;  
Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLALVGKTACGFSLSLLESLSLDPDWTBDQVDSYEDYNOENTSSLT 60  
Db 1 MFARMSDLHVLMLALVGKTACGFSLSLLESLSLDPDWTBDQVDSYEDYNOENTSSLT 60  
QY 61 HAENPDWYVTEQADPCQPNPCFCEHGGDCLVHGSTFTTCSCLAPPSGNKCKQVQNTCKDNPC 120  
Db 61 HAENPDWYVTEQADPCQPNPCFCEHGGDCLVHGSTFTTCSCLAPPSGNKCKQVQNTCKDNPC 120  
QY 121 GRGQCLITQSPPPYRCVCKHPYTGPCSCQVVPVCRPNPCONGATCSRHKERSKFTCACPD 180  
Db 121 GRGQCLITQSPPPYRCVCKHPYTGPCSCQVVPVCRPNPCONGATCSRHKERSKFTCACPD 180  
QY 181 QFKGKFCFCEISDDCYVGDGYSYRGKMNRTVNHQACLYWNSHLLLOENYNNFMEDAEATHGI 240  
Db 181 QFKGKFCFCEISDDCYVGDGYSYRGKMNRTVNHQACLYWNSHLLLOENYNNFMEDAEATHGI 240  
QY 241 GEHNFCEPNDADEKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300  
Db 241 GEHNFCEPNDADEKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300  
QY 301 CGKTEIAERKIKRIYGGFKSTAGHPQWASLQSSPLTISMPQGHFCGGALIHPCWVLTA 360  
Db 301 CGKTEIAERKIKRIYGGFKSTAGHPQWASLQSSPLTISMPQGHFCGGALIHPCWVLTA 360  
QY 361 AHCTDIKTRHLKVVLDQDLKKEEFHQSPRVEKIFKYSHYNERDEIPHNDIALLLKLPV 420  
Db 361 AHCTDIKTRHLKVVLDQDLKKEEFHQSPRVEKIFKYSHYNERDEIPHNDIALLLKLPV 420  
QY 421 DGHCALESKVKVTCLPDGSPFSGSECHISGWGTETGKSRQLLDAKVKLIANTLCNSR 480  
Db 421 DGHCALESKVKVTCLPDGSPFSGSECHISGWGTETGKSRQLLDAKVKLIANTLCNSR 480  
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCEKDGSGPLTCEKDGYYVYGVISWGLECGKRPVY 540  
Db 481 QLYDHMIDDSMICAGNLQKPGQDTCEKDGSGPLTCEKDGYYVYGVISWGLECGKRPVY 540  
QY 541 TQVTKFLNWKATIKSESSEGF 560  
Db 541 TQVTKFLNWKATIKSESSEGF 560

RESULT 2  
US-10-172-712-32  
; Sequence 32, Application US/1017712  
; Publication No. US2003012523A1  
; GENERAL INFORMATION:  
; APPLICANT: GRIFFIN, JOHN H.  
; APPLICANT: GALE, ANDREW J.  
; APPLICANT: PELLETIER, JEAN-LUC  
; APPLICANT: GETZOFF, ELIZABETH D.  
; TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS  
; FILE REFERENCE: 4198-4001US1  
; CURRENT APPLICATION NUMBER: US/10/172,712  
; PRIOR FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: 60/298,578  
; PRIOR FILING DATE: 2001-05-14  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 32  
; LENGTH: 560  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-172-712-32

Query Match 100.0%; Score 3154; DB 14; Length 560;  
Best Local Similarity 100.0%; Pred. No. 9.1e-241;  
Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLALVGKTACGFSLSLLESLSLDPDWTBDQVDSYEDYNOENTSSLT 60  
Db 1 MFARMSDLHVLMLALVGKTACGFSLSLLESLSLDPDWTBDQVDSYEDYNOENTSSLT 60  
QY 61 HAENPDWYVTEQADPCQPNPCFCEHGGDCLVHGSTFTTCSCLAPPSGNKCKQVQNTCKDNPC 120

Db 61 HAENPDWYVTEQADPCQPNPCFCEHGGDCLVHGSTFTTCSCLAPPSGNKCKQVQNTCKDNPC 120  
QY 121 GRGQCLITQSPPPYRCVCKHPYTGPCSCQVVPVCRPNPCONGATCSRHKERSKFTCACPD 180  
Db 121 GRGQCLITQSPPPYRCVCKHPYTGPCSCQVVPVCRPNPCONGATCSRHKERSKFTCACPD 180  
QY 181 QFKGKFCFCEISDDCYVGDGYSYRGKMNRTVNHQACLYWNSHLLLOENYNNFMEDAEATHGI 240  
Db 181 QFKGKFCFCEISDDCYVGDGYSYRGKMNRTVNHQACLYWNSHLLLOENYNNFMEDAEATHGI 240  
QY 241 GEHNFCEPNDADEKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300  
Db 241 GEHNFCEPNDADEKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300  
QY 301 CGKTEIAERKIKRIYGGFKSTAGHPQWASLQSSPLTISMPQGHFCGGALIHPCWVLTA 360  
Db 301 CGKTEIAERKIKRIYGGFKSTAGHPQWASLQSSPLTISMPQGHFCGGALIHPCWVLTA 360  
QY 361 AHCTDIKTRHLKVVLDQDLKKEEFHQSPRVEKIFKYSHYNERDEIPHNDIALLLKLPV 420  
Db 361 AHCTDIKTRHLKVVLDQDLKKEEFHQSPRVEKIFKYSHYNERDEIPHNDIALLLKLPV 420  
QY 421 DGHCALESKVKVTCLPDGSPFSGSECHISGWGTETGKSRQLLDAKVKLIANTLCNSR 480  
Db 421 DGHCALESKVKVTCLPDGSPFSGSECHISGWGTETGKSRQLLDAKVKLIANTLCNSR 480  
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCEKDGSGPLTCEKDGYYVYGVISWGLECGKRPVY 540  
Db 481 QLYDHMIDDSMICAGNLQKPGQDTCEKDGSGPLTCEKDGYYVYGVISWGLECGKRPVY 540  
QY 541 TQVTKFLNWKATIKSESSEGF 560  
Db 541 TQVTKFLNWKATIKSESSEGF 560

RESULT 3  
US-10-391-215-5  
; Sequence 5, Application US/10391215  
; Publication No. US20040009543A1  
; GENERAL INFORMATION:  
; APPLICANT: KIEHL, STEFAN  
; APPLICANT: WILLEIT, JOHANN  
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF  
; APPLICANT: ROEMISCH, JUERGEN  
; APPLICANT: WEIMER, THOMAS  
; APPLICANT: FEUSSNER, ANNETTE  
; APPLICANT: STOEHR, HANS-ARNOLD  
; APPLICANT: DOERSAM, VOLKER  
; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE  
; FILE REFERENCE: 06478.1457-01  
; CURRENT APPLICATION NUMBER: US/10/391,215  
; PRIOR FILING DATE: 2003-03-19  
; PRIOR APPLICATION NUMBER: 09/912,559  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: DE 100 36 641.4  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: DE 100 50 040.4  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: DE 100 52 319.6  
; PRIOR FILING DATE: 2000-10-21  
; PRIOR APPLICATION NUMBER: DE 101 18 706.8  
; PRIOR FILING DATE: 2001-04-12  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 560  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-391-215-5

Query Match 100.0%; Score 3154; DB 15; Length 560;

```
Best Local Similarity 100.0%; Pred. No. 9.1e-241;
Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLALVGTACGFSLSLLESLEDPDWTDPQDYDYSYEDYNQENETSSTLT 60
Db 1 MFARMSDLHVLMLALVGTACGFSLSLLESLEDPDWTDPQDYDYSYEDYNQENETSSTLT 60
QY 61 HAENPDWYVTEQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKCKQVQNTCKDNPC 120
Db 61 HAENPDWYVTEQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKCKQVQNTCKDNPC 120
QY 121 GRGCLITQSPPYRVCVKHPYTGPSCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180
Db 121 GRGCLITQSPPYRVCVKHPYTGPSCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180
QY 181 QFKGKFCBEGSDDDCYVGDGYSGRGNRTVNHACLYWNSHLLLOENYMFMEDEATHGI 240
Db 181 QFKGKFCBEGSDDDCYVGDGYSGRGNRTVNHACLYWNSHLLLOENYMFMEDEATHGI 240
QY 241 GEHNFCRNPDADKPCWCFIKVTNDKVWYCDVSACSADQVAYPEESPTPESTKLPGFDS 300
Db 241 GEHNFCRNPDADKPCWCFIKVTNDKVWYCDVSACSADQVAYPEESPTPESTKLPGFDS 300
QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSSPLTISMPQGHFCGGLIHPCWVLTA 360
Db 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSSPLTISMPQGHFCGGLIHPCWVLTA 360
QY 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSFVKEIFKYSHYNERDEIPHNDIALLLKPKV 420
Db 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSFVKEIFKYSHYNERDEIPHNDIALLLKPKV 420
QY 421 DGHCALESKYVKTVCPLPDGSPSGSECHISGMWVTEGKSRQLLDKAKVLIANTLNSR 480
Db 421 DGHCALESKYVKTVCPLPDGSPSGSECHISGMWVTEGKSRQLLDKAKVLIANTLNSR 480
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCQDGGPLTCEKDGTYVYGVISWGLECGKRPVY 540
Db 481 QLYDHMIDDSMICAGNLQKPGQDTCQDGGPLTCEKDGTYVYGVISWGLECGKRPVY 540
QY 541 TQVTKFLNWKATIKSES GF 560
Db 541 TQVTKFLNWKATIKSES GF 560

RESULT 4
US-10-391-215-7
; Sequence 7, Application US/10391215
; Publication No. US20040009543A1
; GENERAL INFORMATION:
; APPLICANT: KIECHL, STEFAN
; APPLICANT: WILLEIT, JOHANN
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: WEIMER, THOMAS
; APPLICANT: FEUSSNER, ANNETTE
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: DOERSAM, VOLKER
; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE
; FILE REFERENCE: 06478.1457-01
; CURRENT FILING DATE: 2003-03-19
; PRIOR APPLICATION NUMBER: US/10/391.215
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: DE 100 50 040.4
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: DE 100 52 319.6
; PRIOR FILING DATE: 2000-10-21
; PRIOR APPLICATION NUMBER: DE 101 18 706.8
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-391-215-7

Query Match 99.9%; Score 3151; DB 15; Length 560;
Best Local Similarity 99.8%; Pred. No. 1.6e-240;
Matches 559; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLALVGTACGFSLSLLESLEDPDWTDPQDYDYSYEDYNQENETSSTLT 60
Db 1 MFARMSDLHVLMLALVGTACGFSLSLLESLEDPDWTDPQDYDYSYEDYNQENETSSTLT 60
QY 61 HAENPDWYVTEQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKCKQVQNTCKDNPC 120
Db 61 HAENPDWYVTEQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKCKQVQNTCKDNPC 120
QY 121 GRGCLITQSPPYRVCVKHPYTGPSCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180
Db 121 GRGCLITQSPPYRVCVKHPYTGPSCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180
QY 181 QFKGKFCBEGSDDDCYVGDGYSGRGNRTVNHACLYWNSHLLLOENYMFMEDEATHGI 240
Db 181 QFKGKFCBEGSDDDCYVGDGYSGRGNRTVNHACLYWNSHLLLOENYMFMEDEATHGI 240
QY 241 GEHNFCRNPDADKPCWCFIKVTNDKVWYCDVSACSADQVAYPEESPTPESTKLPGFDS 300
Db 241 GEHNFCRNPDADKPCWCFIKVTNDKVWYCDVSACSADQVAYPEESPTPESTKLPGFDS 300
QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSSPLTISMPQGHFCGGLIHPCWVLTA 360
Db 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSSPLTISMPQGHFCGGLIHPCWVLTA 360
QY 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSFVKEIFKYSHYNERDEIPHNDIALLLKPKV 420
Db 361 AHCTDIKTRHLKVVLGDQDLKKEEFHQSFVKEIFKYSHYNERDEIPHNDIALLLKPKV 420
QY 421 DGHCALESKYVKTVCPLPDGSPSGSECHISGMWVTEGKSRQLLDKAKVLIANTLNSR 480
Db 421 DGHCALESKYVKTVCPLPDGSPSGSECHISGMWVTEGKSRQLLDKAKVLIANTLNSR 480
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCQDGGPLTCEKDGTYVYGVISWGLECGKRPVY 540
Db 481 QLYDHMIDDSMICAGNLQKPGQDTCQDGGPLTCEKDGTYVYGVISWGLECGKRPVY 540
QY 541 TQVTKFLNWKATIKSES GF 560
Db 541 TQVTKFLNWKATIKSES GF 560

RESULT 5
US-10-391-215-6
; Sequence 6, Application US/10391215
; Publication No. US20040009543A1
; GENERAL INFORMATION:
; APPLICANT: KIECHL, STEFAN
; APPLICANT: WILLEIT, JOHANN
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF
; APPLICANT: ROEMISCH, JUERGEN
; APPLICANT: WEIMER, THOMAS
; APPLICANT: FEUSSNER, ANNETTE
; APPLICANT: STOEHR, HANS-ARNOLD
; APPLICANT: DOERSAM, VOLKER
; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE
; FILE REFERENCE: 06478.1457-01
; CURRENT FILING DATE: 2003-03-19
; PRIOR APPLICATION NUMBER: US/10/391.215
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: DE 100 36 641.4
```

;; PRIOR FILING DATE: 2000-07-26  
;; PRIOR APPLICATION NUMBER: DE 100 50 040.4  
;; PRIOR FILING DATE: 2000-10-10  
;; PRIOR APPLICATION NUMBER: DE 100 52 319.6  
;; PRIOR FILING DATE: 2000-10-21  
;; PRIOR APPLICATION NUMBER: DE 101 18 706.8  
;; PRIOR FILING DATE: 2001-04-12  
;; NUMBER OF SEQ ID NOS: 8  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 6  
;; LENGTH: 560  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-391-215-6

Query Match 99.7%; Score 3146; DB 15; Length 560;  
Best Local Similarity 99.8%; Pred. No. 3.9e-240;  
Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MFARMSDLHVLNMAVGKTACGFSLSLLES	DPDWT	PDQYDYSYEDYNOE	NTS	STLT	60
Db	1	MFARMSDLHVLNMAVGKTACGFSLSLLES	DPDWT	PDQYDYSYEDYNOE	NTS	STLT	60
Qy	61	HAENPDWYTTEDQADPCQNPCEHGGDCLVHGS	FTTCS	CLAPFSGNKKQKQVNTCKDNPC	120		
Db	61	HAENPDWYTTEDQADPCQNPCEHGGDCLVHGS	FTTCS	CLAPFSGNKKQKQVNTCKDNPC	120		
Qy	121	GRGCLITQSPPYRVCVKHPYTCPSQVVPVCRPNPCNGATCSRHKRSKFTACPD	180				
Db	121	GRGCLITQSPPYRVCVKHPYTCPSQVVPVCRPNPCNGATCSRHKRSKFTACPD	180				
Qy	181	QFKGKFCIGSDDCVVGDGYSYRGKQNRVTNQHACLYWNSHLLQENYNNMFMEDAE	THGI	240			
Db	181	QFKGKFCIGSDDCVVGDGYSYRGKQNRVTNQHACLYWNSHLLQENYNNMFMEDAE	THGI	240			
Qy	241	GEHNFRCNPDADEKPCWCFIKVNDKVWEYCDVSACSAQDVAYPEES	TEPSTKLP	GFDS	300		
Db	241	GEHNFRCNPDADEKPCWCFIKVNDKVWEYCDVSACSAQDVAYPEES	TEPSTKLP	GFDS	300		
Qy	301	CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLP	LTISMPQGH	FCGGALIHPCWV	LT	360	
Db	301	CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLP	LTISMPQGH	FCGGALIHPCWV	LT	360	
Qy	361	AHCTDITKTRHLKVVLGDQDLKKEEFHEQS	FRVEKIFKYS	HYNERDEIPHNDIAL	KLKPV	420	
Db	361	AHCTDITKTRHLKVVLGDQDLKKEEFHEQS	FRVEKIFKYS	HYNERDEIPHNDIAL	KLKPV	420	
Qy	421	DGHCALESKYVKTCLPDGSPFSGSECHISGWGVTETGKSRQLD	DAKVL	LIANTL	CNSR	480	
Db	421	DGHCALESKYVKTCLPDGSPFSGSECHISGWGVTETGKSRQLD	DAKVL	LIANTL	CNSR	480	
Qy	481	QLYDHMIDDSMTCAGNLQKPGQDTCCGDSGGPLTCEKDG	TYVYVYGI	VS	WGLE	CEKRP	GVY 540
Db	481	QLYDHMIDDSMTCAGNLQKPGQDTCCGDSGGPLTCEKDG	TYVYVYGI	VS	WGLE	CEKRP	GVY 540
Qy	541	TQVTKFLNWKATIKSES	SGF	560			
Db	541	TQVTKFLNWKATIKSES	SGF	560			

RESULT 6  
US-09-912-559-4  
; Sequence 4, Application US/09912559  
; Patent No. US20020142316A1  
; GENERAL INFORMATION:  
; APPLICANT: ROEMISCH, JUERGEN  
; APPLICANT: STOEHR, HANS-ARNOLD  
; APPLICANT: FEUSSNER, ANNETTE  
; APPLICANT: LANG, WIEGAND  
; APPLICANT: WEIMER, THOMAS  
; APPLICANT: BECKER, MARGRET  
; APPLICANT: NERLICH, CLAUDIA  
; APPLICANT: MUTH-NAUMANN, GUDRUN

;; TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND  
;; TITLE OF INVENTION: DETECTION METHODS USING SPECIFIC ANTIBODIES  
;; FILE REFERENCE: 06478.1457  
;; CURRENT APPLICATION NUMBER: US/09/912.559  
;; CURRENT FILING DATE: 2001-07-26  
;; PRIOR APPLICATION NUMBER: DE 100 36 641.4  
;; PRIOR FILING DATE: 2000-07-26  
;; PRIOR APPLICATION NUMBER: DE 100 50 040.4  
;; PRIOR FILING DATE: 2000-10-10  
;; PRIOR APPLICATION NUMBER: DE 100 52 319.6  
;; PRIOR FILING DATE: 2000-10-21  
;; PRIOR APPLICATION NUMBER: DE 101 18 706.8  
;; PRIOR FILING DATE: 2001-04-12  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 4  
;; LENGTH: 560  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-912-559-4

Query Match 99.7%; Score 3143; DB 9; Length 560;  
Best Local Similarity 99.6%; Pred. No. 6.7e-240;  
Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MFARMSDLHVLNMAVGKTACGFSLSLLES	DPDWT	PDQYDYSYEDYNOE	NTS	STLT	60
Db	1	MFARMSDLHVLNMAVGKTACGFSLSLLES	DPDWT	PDQYDYSYEDYNOE	NTS	STLT	60
Qy	61	HAENPDWYTTEDQADPCQNPCEHGGDCLVHGS	FTTCS	CLAPFSGNKKQKQVNTCKDNPC	120		
Db	61	HAENPDWYTTEDQADPCQNPCEHGGDCLVHGS	FTTCS	CLAPFSGNKKQKQVNTCKDNPC	120		
Qy	121	GRGCLITQSPPYRVCVKHPYTCPSQVVPVCRPNPCNGATCSRHKRSKFTACPD	180				
Db	121	GRGCLITQSPPYRVCVKHPYTCPSQVVPVCRPNPCNGATCSRHKRSKFTACPD	180				
Qy	181	QFKGKFCIGSDDCVVGDGYSYRGKQNRVTNQHACLYWNSHLLQENYNNMFMEDAE	THGI	240			
Db	181	QFKGKFCIGSDDCVVGDGYSYRGKQNRVTNQHACLYWNSHLLQENYNNMFMEDAE	THGI	240			
Qy	241	GEHNFRCNPDADEKPCWCFIKVNDKVWEYCDVSACSAQDVAYPEES	TEPSTKLP	GFDS	300		
Db	241	GEHNFRCNPDADEKPCWCFIKVNDKVWEYCDVSACSAQDVAYPEES	TEPSTKLP	GFDS	300		
Qy	301	CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLP	LTISMPQGH	FCGGALIHPCWV	LT	360	
Db	301	CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLP	LTISMPQGH	FCGGALIHPCWV	LT	360	
Qy	361	AHCTDITKTRHLKVVLGDQDLKKEEFHEQS	FRVEKIFKYS	HYNERDEIPHNDIAL	KLKPV	420	
Db	361	AHCTDITKTRHLKVVLGDQDLKKEEFHEQS	FRVEKIFKYS	HYNERDEIPHNDIAL	KLKPV	420	
Qy	421	DGHCALESKYVKTCLPDGSPFSGSECHISGWGVTETGKSRQLD	DAKVL	LIANTL	CNSR	480	
Db	421	DGHCALESKYVKTCLPDGSPFSGSECHISGWGVTETGKSRQLD	DAKVL	LIANTL	CNSR	480	
Qy	481	QLYDHMIDDSMTCAGNLQKPGQDTCCGDSGGPLTCEKDG	TYVYVYGI	VS	WGLE	CEKRP	GVY 540
Db	481	QLYDHMIDDSMTCAGNLQKPGQDTCCGDSGGPLTCEKDG	TYVYVYGI	VS	WGLE	CEKRP	GVY 540
Qy	541	TQVTKFLNWKATIKSES	SGF	560			
Db	541	TQVTKFLNWKATIKSES	SGF	560			

RESULT 7  
US-10-391-215-8  
; Sequence 8, Application US/10391215  
; Publication No. US20040009543A1  
; GENERAL INFORMATION:  
; APPLICANT: KIECHL, STEFAN  
; APPLICANT: WILLEIT, JOHANN

APPLICANT: WIEDERMANN, CHRISTIAN JOSEF  
APPLICANT: ROEMISCH, JUERGEN  
APPLICANT: WEIMER, THOMAS  
APPLICANT: FEUSNER, ANNETTE  
APPLICANT: STOEHR, HANS-ARNOLD  
APPLICANT: DOERSAM, VOLKER  
TITLE OF INVENTION: NARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE  
FILE REFERENCE: 06478.1457-01  
CURRENT APPLICATION NUMBER: US/10/391,215  
CURRENT FILING DATE: 2003-03-19  
PRIORITY APPLICATION NUMBER: 09/912,559  
PRIORITY FILING DATE: 2001-07-26  
PRIORITY FILING DATE: 2000-07-26  
PRIORITY FILING DATE: 2000-10-10  
PRIORITY FILING DATE: 2000-10-10  
PRIORITY FILING DATE: 2000-10-21  
PRIORITY FILING DATE: 2001-04-12  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: Patent in Ver. 2.1  
LENGTH: 560  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-391-215-8

Query Match 99.7%; Score 3143; DB 15; Length 560;  
Best Local Similarity 99.6%; Pred. No. 6.7e-240;  
Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLALVGTAGCFSLMSLESILDPTDQDYSDYEDVNOENTSSLT 60  
DB 1 MFARMSDLHVLMLALVGTAGCFSLMSLESILDPTDQDYSDYEDVNOENTSSLT 60

QY 61 HAENPDWYTEDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFGSKGKQKQVONTCKDNPC 120  
DB 61 HAENPDWYTEDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFGSKGKQKQVONTCKDNPC 120

QY 121 GRGCLITOSPPIYRCVCKHPTGSCSQVPCPNPCQNGATCSRHKRSKFTACPD 180  
DB 121 GRGCLITOSPPIYRCVCKHPTGSCSQVPCPNPCQNGATCSRHKRSKFTACPD 180

QY 181 QFKGKFCETGSDDCVYGVGDSYRGKMRNVNOHACLYMNSHLLLOENYNNMEDAETHGI 240  
DB 181 QFKGKFCETGSDDCVYGVGDSYRGKMRNVNOHACLYMNSHLLLOENYNNMEDAETHGI 240

QY 241 GEHNFRCNPDADKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEESPTSTKLPFGDS 300  
DB 241 GEHNFRCNPDADKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEESPTSTKLPFGDS 300

QY 301 CGKTEAERKIKRIYGGFKSTAGKHPQWASLSSPLTISMPQGHPCGALLHPCWVLT 360  
DB 301 CGKTEAERKIKRIYGGFKSTAGKHPQWASLSSPLTISMPQGHPCGALLHPCWVLT 360

QY 361 AHCTDIKTRHLKVLGDQDLKKEEFHQSFVEKIFKYSHYNERDBIPNDIALLKRPV 420  
DB 361 AHCTDIKTRHLKVLGDQDLKKEEFHQSFVEKIFKYSHYNERDBIPNDIALLKRPV 420

QY 421 DGHCALESKYVTKVCLPDGSPFGSGEHSI SGWGTETGKSGRLDADKVLANTLCNSR 480  
DB 421 DGHCALESKYVTKVCLPDGSPFGSGEHSI SGWGTETGKSGRLDADKVLANTLCNSR 480

QY 481 QLYDHMIDDSMI CAGNLQKPGQDTCQDGGGGLTCCKDGTYYYGVIVSGLGCKRPGVY 540  
DB 481 QLYDHMIDDSMI CAGNLQKPGQDTCQDGGGGLTCCKDGTYYYGVIVSGLGCKRPGVY 540

QY 541 TQVTKFLNWKATIKSESGF 560  
DB 541 TQVTKFLNWKATIKSESGF 560

RESULT 8

US-10-172-712-28  
Sequence 28, Application US/10172712  
Publication No. US20030125232A1  
GENERAL INFORMATION:  
APPLICANT: GRIFFIN, JOHN H.  
APPLICANT: GALE, ANDREW J.  
APPLICANT: GETZOFF, ELIZABETH D.  
APPLICANT: PELLEQUER, JEAN-LUC  
TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS  
FILE REFERENCE: 4198-4001US1  
CURRENT APPLICATION NUMBER: US/10/172,712  
CURRENT FILING DATE: 2002-09-30  
PRIORITY APPLICATION NUMBER: 60/298,578  
PRIORITY FILING DATE: 2001-06-14  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 28  
LENGTH: 655  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-172-712-28

Query Match 26.4%; Score 832.5; DB 14; Length 655;  
Best Local Similarity 33.2%; Pred. No. 3.9e-57;  
Matches 192; Conservative 87; Mismatches 216; Indels 83; Gaps 21;

QY 31 ESLDPDWTPDQDYSDYEDVNOENTSSLTTHAENPDW-----YVTEQDA----- 74  
DB 99 QALTEGRPCRFPRYGRGRLHACTSEGSAHRK---WCATTHYDRDRAWGYCVAETPPP 155

QY 75 -----DPCQNPCEHGGDCLVHGSTFTCSCLAPFGSKGKQKQVONTCKDNPCGRGQCL 126  
DB 156 GGPAALDPCASGCLNGGSCNTQDPQSYHCSCPRAFTG-----KD--CGTEKCF 203

QY 127 ITGSPPIY-----RCVCKHPTGSCSQV--PVCNPNPCQNGATCSRHK 169  
DB 204 DETRYELEGDRWARVRQGHVEQCEC---FGRTWCEGTRHTACUSSPCLNGGTCHLIV 260

QY 170 RRSKFTCACPDQKFKCEIGSD--CVVGDYSGYRGKMRNVNOHACLYMNSHLLLOENY 228  
DB 261 ATGTTVCACPPGAGRLCNIEPDERCFLNGTGYRGVASTSASGLSCLAWNSDLYQELH 320

QY 229 NMFMEDAEHGI GBHNFRCNPDADKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEE 286  
DB 321 VDSVGAALLGLGPHAYCRNPNDNERPCYV--VKDSALSWEYCRLEACESLTRVQLSPDL 379

QY 287 SPTSPSTKLPQFSCGKTEAERKIK--RIYGGFKSTAGKHPQWASLSSPLTISMPQGH 345  
DB 380 LATPEPASFORQACGRHKKRTFLRPRIIGSSSLPGSHPLAAIYG-----DS 430

QY 346 FCGGALLHPCWVLTAAHCTDIKTRH--LKVLGDQDLKKEEFHQSFVEKIFKYSHYNE 403  
DB 431 FCAGSLVHTCWVYSAACHFSHPDRSVSVLQGHFFNRTTDTVTQFGIEKIPYTLYSV 490

QY 404 RDETPHNDIALLKLPVDGHCALESKYVTKVCLPD--GSPPSGSECHISGWW-VTETGK 460  
DB 491 FNPSDH--DLVILRLKKGDRCATRSQVQVICIPEGSGTFPAGHKQCIAGHLDENVSG 549

QY 461 -SRQLLDKAVKLANTLCNSRQLYDHMIDDSMI CAGNLQKPGQDTCQDGGGGLTCCKDG 519  
DB 550 YSSSLREALVPLVADHKCSSPEVYGADISPNMLCAGYFDCK--SDACQDGGGGLACEKNG 608

QY 520 TYYYGVIVSGLGCKG--RFGVTVQVTKFLNWKATIK 555  
DB 609 VAVLYGIISGDDGGCGRLEKFGVTVRVANYVDWINDIR 646

RESULT 9

US-09-974-298-145  
Sequence 145, Application US/09974298  
Patent No. US20020156263A1

GENERAL INFORMATION:  
APPLICANT: Chen, Hui-Mei  
TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER  
FILE REFERENCE: PA-0037 P  
CURRENT APPLICATION NUMBER: US/09/974,299  
CURRENT FILING DATE: 2001-10-04  
PRIOR APPLICATION NUMBER: 60/238,331  
PRIOR FILING DATE: 2000-05-10  
NUMBER OF SEQ ID NOS: 194  
SOFTWARE: PERL Program  
SEQ ID NO 145  
LENGTH: 562  
TYPE: PRT  
ORGANISM: Homo sapiens  
NAME/KEY: misc feature  
OTHER INFORMATION: incyte ID No. US20020156263A1 1001470CD1  
US-09-974-298-145

Query Match 24.3%; Score 767.5; DB 9; Length 562;  
Best Local Similarity 35.0%; Pred. No. 4.4e-52;  
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG-----123  
Db 86 CSEPRCFNGGTCQOALYFSDVCCPEGFAGKCCCEIDTRATCYEDQGISYRGTWSTAESG 145  
QY 124 -QCLITOSPYPYRCVKHPYTPGSCSQVVPVCRPNPCQ----NGATCSRHRKRSKFTCAC 178  
Db 146 AECTWNNS----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDRLDRSKPCVY 192  
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMNRTVNOHACLWNHLLQENY 228  
Db 193 ---FRAGKYSSEFCSTPACSEGNDCYFGNSAYRGTHSLTESGASCLPWNMILIGKVY 249  
QY 229 NMFMEDATHGIGEHNFORNPDADKPCWCFIKVTNDKWKWEYCDVSACSADQVAYPEESP 288  
Db 250 TAQNPQAQALGLGKHNYCRNPDGAKPWCHV-LXNRLTLWEYCDVPSCS-----297  
QY 289 TEPSTKLPGFDSGCKTEIAERKIKRIYGGFKSTAGKHPQASLOSLLPLTISMPQGHFCG 348  
Db 298 -----TCGLRQYSQQF-RIKGGLFADIASHPWQAIAFAK--HRRSPGERFLCG 343  
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVLGQDLKKEEFHQSPRVEKIFKYSHYNERDE 406  
Db 344 GILISSCWILSAACFCQERFPFPHLTVILGRTYRVVPGEEOQKEVEKYIVHKEFD--DD 401  
QY 407 IPHNDIALKLPVDGCHCALESKYVKTVCPLP--DGSPSGSECHISGMYETCTGK--SR 462  
Db 402 TYDNDIALQLKSDSSRCAQESSVVRTVCLPADLQDPDTECELSGYGKEALSPPYSE 461  
QY 463 QLLDAKVLIANTLNSRLQYDHMIDDSMI CAGNLQKPG-----QDTCQSDSGGFLTCEK 517  
Db 462 RLKEAHVRLYPSRRTCSQHLNRTVTDNMLCAGDTRSGGPOANLHDACQSDSGGFLVCLN 521  
QY 518 DGTYYVYGVSVGLECGKR--PGVYTVTKFLNWKATIK 555  
Db 522 DGRMTLVGIISGLGCGQKDVGVYTKVTNYLDWIRDNR 561

RESULT 10  
US-09-987-457-18  
Sequence 18, Application US/09987457  
Publication No. US20030013150A1  
GENERAL INFORMATION:  
APPLICANT: Manosroi, Aranya  
APPLICANT: Manosroi, Jiradej  
APPLICANT: Tavapiwatana, Chatchai  
APPLICANT: Goetz, Friedrich  
APPLICANT: Werner, Rolf-Guenther  
TITLE OF INVENTION: Methods for Large Scale Protein Production in Prokaryotes  
FILE REFERENCE: 0652.2180001  
CURRENT APPLICATION NUMBER: US/09/987,457

GENERAL INFORMATION:  
APPLICANT: Chen, Hui-Mei  
TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER  
FILE REFERENCE: PA-0037 P  
CURRENT APPLICATION NUMBER: US/09/974,299  
CURRENT FILING DATE: 2001-10-04  
PRIOR APPLICATION NUMBER: 60/238,331  
PRIOR FILING DATE: 2000-05-10  
NUMBER OF SEQ ID NOS: 194  
SOFTWARE: PERL Program  
SEQ ID NO 145  
LENGTH: 562  
TYPE: PRT  
ORGANISM: Homo sapiens  
NAME/KEY: misc feature  
OTHER INFORMATION: incyte ID No. US20020156263A1 1001470CD1  
US-09-974-298-145

Query Match 24.3%; Score 766.5; DB 10; Length 527;  
Best Local Similarity 35.0%; Pred. No. 4.9e-52;  
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG-----123  
Db 51 CSEPRCFNGGTCQOALYFSDVCCPEGFAGKCCCEIDTRATCYEDQGISYRGTWSTAESG 110  
QY 124 -QCLITOSPYPYRCVKHPYTPGSCSQVVPVCRPNPCQ----NGATCSRHRKRSKFTCAC 178  
Db 111 AECTWNNS----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDRLDRSKPCVY 157  
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMNRTVNOHACLWNHLLQENY 228  
Db 158 ---FRAGKYSSEFCSTPACSEGNDCYFGNSAYRGTHSLTESGASCLPWNMILIGKVY 214  
QY 229 NMFMEDATHGIGEHNFORNPDADKPCWCFIKVTNDKWKWEYCDVSACSADQVAYPEESP 288  
Db 215 TAQNPQAQALGLGKHNYCRNPDGAKPWCHV-LXNRLTLWEYCDVPSCS-----262  
QY 289 TEPSTKLPGFDSGCKTEIAERKIKRIYGGFKSTAGKHPQASLOSLLPLTISMPQGHFCG 348  
Db 263 -----TCGLRQYSQQF-RIKGGLFADIASHPWQAIAFAK--HRRSPGERFLCG 308  
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVLGQDLKKEEFHQSPRVEKIFKYSHYNERDE 406  
Db 309 GILISSCWILSAACFCQERFPFPHLTVILGRTYRVVPGEEOQKEVEKYIVHKEFD--DD 366  
QY 407 IPHNDIALKLPVDGCHCALESKYVKTVCPLP--DGSPSGSECHISGMYETCTGK--SR 462  
Db 367 TYDNDIALQLKSDSSRCAQESSVVRTVCLPADLQDPDTECELSGYGKEALSPPYSE 426  
QY 463 QLLDAKVLIANTLNSRLQYDHMIDDSMI CAGNLQKPG-----QDTCQSDSGGFLTCEK 517  
Db 427 RLKEAHVRLYPSRRTCSQHLNRTVTDNMLCAGDTRSGGPOANLHDACQSDSGGFLVCLN 486  
QY 518 DGTYYVYGVSVGLECGKR--PGVYTVTKFLNWKATIK 555  
Db 487 DGRMTLVGIISGLGCGQKDVGVYTKVTNYLDWIRDNR 526

RESULT 11  
US-09-987-455-19  
Sequence 19, Application US/09987455  
Publication No. US20030049729A1  
GENERAL INFORMATION:  
APPLICANT: Aranya Manosroi  
APPLICANT: Jiradej Manosroi  
APPLICANT: Chatchai Tavapiwatana  
APPLICANT: Friedrich Goetz  
APPLICANT: Rolf-Guenther Werner  
TITLE OF INVENTION: Methods for Large Scale Production of Recombinant  
FILE REFERENCE: 0652.2190001  
CURRENT APPLICATION NUMBER: US/09/987,455  
CURRENT FILING DATE: 2001-11-14  
PRIOR APPLICATION NUMBER: 60/268,574  
PRIOR FILING DATE: 2001-02-15  
PRIOR APPLICATION NUMBER: GB 0027779.8  
PRIOR FILING DATE: 2000-11-14



; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 527  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-987-455-19

Query Match 24.3%; Score 766.5; DB 10; Length 527;  
Best Local Similarity 35.0%; Pred. No. 4.9e-52;  
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ--KVQNTCKDNP--CGRG-----123  
DB 51 CSEPRCFNGGTCQQALYFSDFCQCEGAGKCCCEIDTRATCYEDQGISYRGTWSTAESG 110  
QY 124 -QCLITQSPYRVCVKHPYTGPSQSVVPCRNPCQ-----NGATCSRHKRSKFTCAC 178  
DB 111 AECTWNWS-----SALAQKPYSGR-----RPAIRLGLGNHNYCRNPDSDSKPMCYV 157  
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMNRTVNOHACLYWNSHLLQENY 228  
DB 158 ---FKAGKYSSEFCSTPACSEGNSDCYFGNSAYRGTHSLTESGASCLPWNMILIGKYV 214  
QY 229 NMFEDEATHGIEHNFCRNPDADKPMCFIKVTNDKVKWEYCDVSACSADQVAYPEESP 288  
DB 215 TAQNPQAQALGLGKHNKYNCRNPDGDAKPMCHV--LKNRLTWECYCDVPSCS-----262  
QY 289 TEPSTKLPGDFSCGCKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCG 348  
DB 263 -----TCGLRQYSQPOF--RIKGLFADIASHPWQAIFAK--HRRSPGERFLCG 308  
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSFVEKIFKYSHYNERDE 406  
DB 309 GILISSCWILSAAHCFQERFPPHLLTVILGRYRVVPGEEQKFEVEKIVHKEFD--DD 366  
QY 407 IPHNDIALKLPVDGHCALSKYKVTCLP--DGSFPGSECHISGWTGTGKG--SR 462  
DB 367 TYNDIALLOKSDSRCAQESSVVRTCLPADLQLPDWTCELSGYGKHEALSPFYSE 426  
QY 463 QLLDAKVLKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----QDTQCGSGGSLTCEK 517  
DB 427 RLKEAHRVLYPSSRCTSQHLLNRTVTDNMLCAGDTRSGGPQANLHDACQDGGSLVCLN 486  
QY 518 DGTYYVYGVISWGLECGKR--PGVYTVTKFLNWKATIK 555  
DB 487 DGRMTLVGIISWGLGCGKQDVPGVYTKVTNYLDWIRDNR 526

RESULT 12  
US-10-432-842-1  
; Sequence 1, Application US/10432842  
; Publication No. US20040071707A1  
; GENERAL INFORMATION:  
; APPLICANT: Veronica A. CARROLL  
; APPLICANT: Adrian L. HARRIS  
; APPLICANT: Roy BICKNELL  
; APPLICANT: Pat PRICE  
; TITLE OF INVENTION: MODULATION OF CELL GROWTH  
; FILE REFERENCE: 117-450 / N.79507A SER  
; CURRENT APPLICATION NUMBER: US/10/432,842  
; CURRENT FILING DATE: 2003-09-27  
; PRIOR APPLICATION NUMBER: PCT/GB01/05244  
; PRIOR FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: GB 0029001.5  
; PRIOR FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: MS Word  
; SEQ ID NO 1  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-432-842-1

Query Match 24.3%; Score 766.5; DB 12; Length 527;  
Best Local Similarity 35.0%; Pred. No. 4.9e-52;  
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ--KVQNTCKDNP--CGRG-----123  
DB 51 CSEPRCFNGGTCQQALYFSDFCQCEGAGKCCCEIDTRATCYEDQGISYRGTWSTAESG 110  
QY 124 -QCLITQSPYRVCVKHPYTGPSQSVVPCRNPCQ-----NGATCSRHKRSKFTCAC 178  
DB 111 AECTWNWS-----SALAQKPYSGR-----RPAIRLGLGNHNYCRNPDSDSKPMCYV 157  
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMNRTVNOHACLYWNSHLLQENY 228  
DB 158 ---FKAGKYSSEFCSTPACSEGNSDCYFGNSAYRGTHSLTESGASCLPWNMILIGKYV 214  
QY 229 NMFEDEATHGIEHNFCRNPDADKPMCFIKVTNDKVKWEYCDVSACSADQVAYPEESP 288  
DB 215 TAQNPQAQALGLGKHNKYNCRNPDGDAKPMCHV--LKNRLTWECYCDVPSCS-----262  
QY 289 TEPSTKLPGDFSCGCKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCG 348  
DB 263 -----TCGLRQYSQPOF--RIKGLFADIASHPWQAIFAK--HRRSPGERFLCG 308  
QY 349 GALIHPCWVLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSFVEKIFKYSHYNERDE 406  
DB 309 GILISSCWILSAAHCFQERFPPHLLTVILGRYRVVPGEEQKFEVEKIVHKEFD--DD 366  
QY 407 IPHNDIALKLPVDGHCALSKYKVTCLP--DGSFPGSECHISGWTGTGKG--SR 462  
DB 367 TYNDIALLOKSDSRCAQESSVVRTCLPADLQLPDWTCELSGYGKHEALSPFYSE 426  
QY 463 QLLDAKVLKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----QDTQCGSGGSLTCEK 517  
DB 427 RLKEAHRVLYPSSRCTSQHLLNRTVTDNMLCAGDTRSGGPQANLHDACQDGGSLVCLN 486  
QY 518 DGTYYVYGVISWGLECGKR--PGVYTVTKFLNWKATIK 555  
DB 487 DGRMTLVGIISWGLGCGKQDVPGVYTKVTNYLDWIRDNR 526

RESULT 13  
US-10-360-101-203  
; Sequence 203, Application US/10360101  
; Publication No. US20040009550A1  
; GENERAL INFORMATION:  
; APPLICANT: Moll, Gert N.  
; APPLICANT: Leenhouts, Cornelis J.  
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way  
; FILE REFERENCE: 2183-5673  
; CURRENT APPLICATION NUMBER: US/10/360,101  
; CURRENT FILING DATE: 2003-02-07  
; PRIOR APPLICATION NUMBER: EP 02077060.8  
; PRIOR FILING DATE: 2002-05-24  
; NUMBER OF SEQ ID NOS: 309  
; SOFTWARE: Patent In version 3.1  
; SEQ ID NO 203  
; LENGTH: 527  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: sequence of alteplase  
US-10-360-101-203

Query Match 24.3%; Score 766.5; DB 15; Length 527;  
Best Local Similarity 35.0%; Pred. No. 4.9e-52;  
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ--KVQNTCKDNP--CGRG-----123  
DB 51 CSEPRCFNGGTCQQALYFSDFCQCEGAGKCCCEIDTRATCYEDQGISYRGTWSTAESG 110

124 -QCLITQSPYYRCVCKHPYTGSPSCSVQVVPVCRPNFCQ-----NGATCSRHKRKRSKFTAC 178  
111 AECTNWS-----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPRDRSKFCV 157  
179 PDQFK-GKF-----CEIGDDCYVGDGYSGYRGKMRNTVNOHACLYWNSHLLLOENY 228  
158 ---FRAGKYSSEFCSTPACSEGNDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKY 214  
229 NMFMEDAETHGIEHNFCRNPDADEKPCWIKVNDKVKWEYCDVGSACSADQVAYPEESP 288  
215 TAQNPSAALGLGKHNVCNPDGAKPCHV-LKNEALTWECYDVFSCS-----262  
289 TEPSTKLPFGDSCGKTEIAERKIRIYGGFKSTAGKHPWQASLOSLPLTISNPOQHFCG 348  
263 -----TCGLRQYSQPF-RIKGGLFADIAHSPWQAIFAK--HRRSPGERFLCG 308  
349 GALHPQWVLTAAHCTD--IKTRHLKVLVGDQDLKKEEFHQSPRVEKIFKYSHYNERDE 406  
309 GILISSCWILSAHCFQERPPHLLTVILGRVYVPGEEQKFEVEKIVHKEFD--DD 366  
407 IPHNDIALKLPVDGHCALSKYKVTCLP--DGSFPGSECHISGWGVTETGKG--SR 462  
367 TYNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGKHEALSPFYSE 426  
463 QLLDAKVKLIANTLCSNRSQLYDHMDSDMICAGNLQKPG-----QDTCQDSDSGPLTCEK 517  
427 RLKEAHVRLYPSRCTQHLNLRVTDNMLCAGDTRSGGPGQANLHDAQDSDSGPLVCLN 486  
518 DGTYYVYGVISWGLECGKR--PGVYTVQTKPLNMIKATIK 555  
487 DGRMTLVGIISWGLCGQKQDVGYYTKVTNYLDWIRDNR 526

RESULT 14  
US-09-969-271-7  
; Sequence 7, Application US/09969271  
; Patent No. US20020098179A1  
; GENERAL INFORMATION:  
; APPLICANT: Pfizer Inc. (All designated States except GB and EP (GB));  
; APPLICANT: Pfizer Limited (GB and EP (GB) only)  
; TITLE OF INVENTION: Pharmaceutical Combinations  
; FILE REFERENCE: PCS10951APME  
; CURRENT APPLICATION NUMBER: US/09/969,271  
; PRIOR FILING DATE: 2001-10-01  
; PRIOR APPLICATION NUMBER: GB 0025473.0  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-969-271-7

Query Match 24.3%; Score 766.5; DB 9; Length 562;  
Best Local Similarity 35.0%; Pred. No. 5.3e-52;  
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;  
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG-----123  
Db 86 CSEPRCFNGGTCCQALYFSDFCVQCPEGFAGKCCCEIDTRATCYEDQGISYRGWTSTAESG 145  
QY 124 -QCLITQSPYYRCVCKHPYTGSPSCSVQVVPVCRPNFCQ-----NGATCSRHKRKRSKFTAC 178  
Db 146 AECTNWS-----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPRDRSKFCV 192  
QY 179 PDQFK-GKF-----CEIGDDCYVGDGYSGYRGKMRNTVNOHACLYWNSHLLLOENY 228  
Db 193 ---FRAGKYSSEFCSTPACSEGNDCYFGNGSAYRGTHSLTESGASCLPWNMILIGKY 249  
QY 229 NMFMEDAETHGIEHNFCRNPDADEKPCWIKVNDKVKWEYCDVGSACSADQVAYPEESP 288  
Db 250 TAQNPSAALGLGKHNVCNPDGAKPCHV-LKNEALTWECYDVFSCS-----297

289 TEPSTKLPFGDSCGKTEIAERKIRIYGGFKSTAGKHPWQASLOSLPLTISNPOQHFCG 348  
298 -----TCGLRQYSQPF-RIKGGLFADIAHSPWQAIFAK--HRRSPGERFLCG 343  
349 GALHPQWVLTAAHCTD--IKTRHLKVLVGDQDLKKEEFHQSPRVEKIFKYSHYNERDE 406  
344 GILISSCWILSAHCFQERPPHLLTVILGRVYVPGEEQKFEVEKIVHKEFD--DD 401  
407 IPHNDIALKLPVDGHCALSKYKVTCLP--DGSFPGSECHISGWGVTETGKG--SR 462  
402 TYNDIALQLKSDSSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGKHEALSPFYSE 461  
463 QLLDAKVKLIANTLCSNRSQLYDHMDSDMICAGNLQKPG-----QDTCQDSDSGPLTCEK 517  
462 RLKEAHVRLYPSRCTQHLNLRVTDNMLCAGDTRSGGPGQANLHDAQDSDSGPLVCLN 521  
518 DGTYYVYGVISWGLECGKR--PGVYTVQTKPLNMIKATIK 555  
522 DGRMTLVGIISWGLCGQKQDVGYYTKVTNYLDWIRDNR 561

RESULT 15  
US-10-411-037-26  
; Sequence 26, Application US/10411037  
; Publication No. US20040043446A1  
; GENERAL INFORMATION:  
; APPLICANT: Neose Technologies, Inc.  
; APPLICANT: DeFees, Shawn  
; APPLICANT: Zopf, David  
; APPLICANT: Bayer, Robert  
; APPLICANT: Hakes, David  
; APPLICANT: Chen, Xi  
; APPLICANT: Bowe, Caryn  
; TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF ALPHA  
; FILE REFERENCE: 040853-01-5082  
; CURRENT APPLICATION NUMBER: US/10/411,037  
; PRIOR FILING DATE: 2003-04-09  
; PRIOR APPLICATION NUMBER: US 60/328,523  
; PRIOR FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 60/344,692  
; PRIOR FILING DATE: 2001-10-19  
; PRIOR APPLICATION NUMBER: US 60/387,292  
; PRIOR FILING DATE: 2002-08-07  
; PRIOR APPLICATION NUMBER: US 60/391,777  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: US 60/396,594  
; PRIOR FILING DATE: 2002-07-17  
; PRIOR APPLICATION NUMBER: US 60/404,249  
; PRIOR APPLICATION NUMBER: US 60/407,527  
; PRIOR FILING DATE: 2002-08-28  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 26  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-411-037-26

Query Match 24.3%; Score 766.5; DB 12; Length 562;  
Best Local Similarity 35.0%; Pred. No. 5.3e-52;  
Matches 182; Conservative 77; Mismatches 176; Indels 85; Gaps 20;  
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG-----123  
Db 86 CSEPRCFNGGTCCQALYFSDFCVQCPEGFAGKCCCEIDTRATCYEDQGISYRGWTSTAESG 145  
QY 124 -QCLITQSPYYRCVCKHPYTGSPSCSVQVVPVCRPNFCQ-----NGATCSRHKRKRSKFTAC 178  
Db 146 AECTNWS-----SALAQKPYSGR-----RPDAIRLGLGNHNYCRNPRDRSKFCV 192

```

179 PDQFK-GKF-----CBIGSDDCVVGCGSYRGKMRNTVQNHACLWYNSHLLQENY 228
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
193 ---FKAGKYSFEFCSTPACSEGNDSYCVFNGSAYRGTHSLTSGASCLEPNWMIILGKVV 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
229 NMFVEDAETHGIGBHNCRMPDADAKPWCIFKVTNDKVKWEYCDVSACSAQDVAYPESSP 288
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
250 TAQNPQAALGLGKHNTCRUPDGDGAKPWCHV-LKNRRLTWIYCDVPSCS----- 297
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
289 TEPSTKLPFGDSCKGTEIAERKIKRIYGGFKSTAGKHPQASLQSSLP-LTISMPQGHFGG 348
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
298 -----TCGLRQYSQPQF-RINGGLFADIASHPQAAIFAK--HRRSPGERFLCG 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
349 GALTHPCWLTPAHC TD--IKTRHLKVLGDQLKKEEFTHEQSFRVEKFKYSHYNERDE 406
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
344 GILISSCWILSAACHFOERFPFPHLTVILGRTYRVVPGEBEEKFYEKIVYHKBFD--DD 401
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
407 IPHNDIALLKLPVDGHCALESKVVKTVCLP--DGSPFSGSECHISGWGVTETGKG--SR 462
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
402 TYNDNDIALLQKSDSSRCABESSVVRIVCLPPADLQLPDWTECELSGYGKHALSPFYSE 461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
463 QLLDAKVLIANTLCNSRQLYDHMDIDSMT-CAGNLQKPG-----QDTQCGDSGGPUTCBK 517
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
462 RLKEAHVRLYFSSRCTSQHLLNRITVDNMLCAGDTRSGGQPQANLHDACQDGGGPLVCLVN 521
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
518 DGTYYVYGIVSWGLECGKR--PGVYTVQTVFLNWKATIK 555
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
522 DGRVTLVGIISWGLGCGCKGVGYVYTKVNYLWDIRDNR 561
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: May 24, 2004, 09:58:57  
Job time : 44 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 09:39:54 ; Search time 54 Seconds  
(without alignments)  
2930.122 Million cell updates/sec

Title: US-09-912-559-4  
Perfect score: 3153  
Sequence: 1 MFARMSDLVLLMALVGKT.....TQVTKFLNLWIKATIKESGF 560

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3153	100.0	560	AA017145	AA017145 Human blo
2	3143	99.7	560	AA024284	AA024284 Human ORF
3	3143	99.7	560	AA017144	AA017144 Human blo
4	822.5	26.1	655	AA053962	AA053962 Hepatocyt
5	821.5	26.1	655	AA089197	AA089197 Human hep
6	801.5	25.4	441	AA070643	AA070643 Modified
7	801.5	25.4	473	AA070642	AA070642 Modified
8	796.5	25.3	689	AA070850	AA070850 Novel pro
9	791	25.1	439	AA09360	AA09360 t-PA vari
10	787.5	25.0	434	AA013922	AA013922 Delta (92
11	786	24.9	439	AA09261	AA09261 t-PA vari
12	784	24.9	439	AA09258	AA09258 t-PA vari
13	783	24.8	439	AA09259	AA09259 t-PA vari
14	782	24.8	623	AA017142	AA017142 Ile(277)t
15	781	24.8	623	AA071740	AA071740 t-PA with
16	775	24.6	527	AA070858	AA070858 Human t-P
17	775	24.6	562	AA041545	AA041545 Modified
18	774	24.5	518	AA041545	AA041545 t-PA mute
19	774	24.5	522	AA013917	AA013917 Delta (46
20	773	24.5	527	AA070893	AA070893 Human t-P
21	772	24.5	521	AA074678	AA074678 t-PA mute
22	771.5	24.5	483	AA09257	AA09257 t-PA vari
23	771	24.5	522	AA074653	AA074653 t-PA mute
24	770.5	24.4	483	AA09252	AA09252 t-PA vari
25	770	24.4	519	AA074657	AA074657 t-PA vari

26	770	24.4	527	2	AA070861	AA070861 Human t-P
27	770	24.4	527	2	AA070856	AA070856 Human t-P
28	770	24.4	527	2	AA070869	AA070869 Human t-P
29	769.5	24.4	483	2	AA09246	AA09246 t-PA vari
30	769	24.4	519	2	AA074680	AA074680 t-PA mute
31	769	24.4	522	2	AA013920	AA013920 Delta (46
32	769	24.4	527	2	AA070874	AA070874 Human t-P
33	769	24.4	527	2	AA070859	AA070859 Human t-P
34	768.5	24.4	483	2	AA070889	AA070889 Human t-P
35	768	24.4	522	2	AA074685	AA074685 t-PA mute
36	768	24.4	527	2	AA070896	AA070896 Human t-P
37	768	24.4	527	2	AA070904	AA070904 Human t-P
38	768	24.4	527	2	AA070891	AA070891 Human t-P
39	767.5	24.3	483	2	AA09250	AA09250 t-PA vari
40	767.5	24.3	483	2	AA09251	AA09251 t-PA vari
41	767.5	24.3	562	2	AA023802	AA023802 Zymogen-1
42	767	24.3	518	2	AA074679	AA074679 t-PA mute
43	767	24.3	522	2	AA014486	AA014486 Delta (46
44	767	24.3	527	2	AA070876	AA070876 Human t-P
45	767	24.3	527	2	AA070865	AA070865 Human t-P

## ALIGNMENTS

RESULT 1  
AA017145  
ID AAO17145 standard; protein; 560 AA.

XX AAO17145;  
XX  
XX 13-JUN-2002 (first entry)  
XX  
XX Human blood coagulation factor VII activating protease mutant.  
XX Human; blood coagulation factor VII activating protease; FSAP;  
XX single-chain plasminogen activator; bleeding disorder; haematological;  
XX haemostatic; mutant; mutein.  
XX Homo sapiens.  
XX Synthetic.  
XX EPI182258-A1.  
XX  
XX 27-FEB-2002.  
XX  
XX 05-JUL-2001; 2001EP-00115691.  
XX  
XX 26-JUL-2000; 2000DE-01036641.  
XX 10-OCT-2000; 2000DE-01050040.  
XX 21-OCT-2000; 2000DE-01052319.  
XX 12-APR-2001; 2001DE-01018706.  
XX (AVET ) AVENTIS BEHRING GMBH.  
XX Roemisch J, Stoehr H, Feussner A, Lang W, Weimer T, Becker M;  
XX Nerlich C, Muth-Naumann G;  
XX WPI; 2002-270939/32.  
XX N-PSDB; AAL45697.  
XX  
XX New nucleic acid encoding mutant factor 7 activating protease, useful for  
XX diagnosis, treatment and prevention of coagulation disorders, also  
XX related protein and antibodies.  
XX Disclosure; Page 20-22; 27pp; German.  
XX  
XX The present invention relates to a mutant of the DNA sequence encoding  
XX the protease (FSAP) that activates blood coagulation factor VII (FVII)  
XX and single-chain plasminogen activator, where at least one of the base  
XX changes G to C at nucleotide 1177 and G to A at nucleotide 1601 is  
XX present. The mutant sequences can be used in the treatment and prevention  
XX of bleeding disorders associated with inherited or acquired defects of

CC blood coagulation factors V, VIII, IX, X, XI, XII, von Willebrand's  
CC factor and/or with antibodies against any of these proteins. The present  
CC sequence is the mutant human FSAP protein  
XX  
SQ Sequence 560 AA;

Query Match 100.0%; Score 3153; DB 5; Length 560;  
Best Local Similarity 100.0%; Pred. No. 7.5e-186; Mismatches 0; Indels 0; Gaps 0;  
Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLVGTACGFSLSMSLSDPDPDQDYDYSDYEDYNOENTSSLT 60  
DB 1 MFARMSDLHVLMLVGTACGFSLSMSLSDPDPDQDYDYSDYEDYNOENTSSLT 60  
QY 61 HAENPDWYTTEDQADPCQPNPCRHGGDCLVHGSTFTCSCLAPFSGNKKQVQNTCKDNP 120  
DB 61 HAENPDWYTTEDQADPCQPNPCRHGGDCLVHGSTFTCSCLAPFSGNKKQVQNTCKDNP 120  
QY 121 GRGQCLITQSPFYRCVCKHPYTGPCSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180  
DB 121 GRGQCLITQSPFYRCVCKHPYTGPCSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180  
QY 181 QFKGKFCIEIGDDCYVGDGYSYRGKWRVTNQHACLYNWSHLLLOENYMFEDAETHGI 240  
DB 181 QFKGKFCIEIGDDCYVGDGYSYRGKWRVTNQHACLYNWSHLLLOENYMFEDAETHGI 240  
QY 241 GEHNFRCNPDADEKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPESPTSTKLPGPS 300  
DB 241 GEHNFRCNPDADEKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPESPTSTKLPGPS 300  
QY 301 CGTEIAERKI KRIYGGFKSTAGKHPQASLOSSLPITISMPQGHFCGALHPCWVLT 360  
DB 301 CGTEIAERKI KRIYGGFKSTAGKHPQASLOSSLPITISMPQGHFCGALHPCWVLT 360  
QY 361 AHCTDIKTRHLKVLGDQDLKKEEFHQSPVQKIFKYSHYNERDEIPHNDIALKLPV 420  
DB 361 AHCTDIKTRHLKVLGDQDLKKEEFHQSPVQKIFKYSHYNERDEIPHNDIALKLPV 420  
QY 421 DGHCALESKYVTKVCLPDGSPFSGSECHI:SGMGVTEGKSGQLDQVLIANTLNSR 480  
DB 421 DGHCALESKYVTKVCLPDGSPFSGSECHI:SGMGVTEGKSGQLDQVLIANTLNSR 480  
QY 481 QLVDMITDDSMICAGNLQKQDQTCGDSGGPLTCBKDTYTYVYGVISWGLECEKRPVY 540  
DB 481 QLVDMITDDSMICAGNLQKQDQTCGDSGGPLTCBKDTYTYVYGVISWGLECEKRPVY 540  
QY 541 TQVTKFLNLWIKATIKSES 560  
DB 541 TQVTKFLNLWIKATIKSES 560

RESULT 2  
AAB42484  
ID AAB42484 standard; protein: 560 AA.  
XX  
AC AAB42484;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF2248 polypeptide sequence SEQ ID NO:4496.

KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX  
OS Homo sapiens.  
XX WC200058473-A2.  
PN  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000; 2000WO-US008621.  
XX  
PP 31-MAR-1999; 99US-0127607P.  
XX  
PR 02-APR-1999; 99US-0127636P.  
PR  
PR 05-APR-1999; 99US-0127728P.  
PR  
PR 30-MAR-2000; 2000US-00540763.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shimkets RA, Leach M;  
XX  
XX WPI; 2000-602362/57.  
XX  
XX N-PSDB; AAC76693.  
XX  
PT Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease.  
XX  
PS Claim 11; Page 3681-3683; 5507pp; English.  
XX  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX  
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;  
XX antiparasitic; antiparkinsonian; nontropic; neuroprotective; osteopathic;  
XX anticovulant; antidiabetic; immunosuppressant; immunostimulant;  
XX cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
XX dermatological; immunosuppressive; antiinflammatory; antibacterial;  
XX antiviral; antifungal; antineumatic; antithyroid; and antianemic. The  
XX sequences can be used for determining the presence of or predisposition  
XX to, or preventing or treating pathological conditions associated with an  
XX ORFX-associated disorder. The nucleic acids can be used to express ORFX  
XX proteins in gene therapy vectors. The proteins and nucleic acids may be  
XX used to treat cancers, proliferative disorders, neurodegenerative  
XX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
XX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
XX storage, systemic lupus erythematosus, severe combined immunodeficiency  
XX (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
XX disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
XX cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
XX enhance coagulation; to inhibit thrombosis; and as a contraceptive

Query Match 99.7%; Score 3143; DB 3; Length 560;  
Best Local Similarity 99.6%; Pred. No. 3.1e-185;  
Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLVGTACGFSLSMSLSDPDPDQDYDYSDYEDYNOENTSSLT 60  
DB 1 MFARMSDLHVLMLVGTACGFSLSMSLSDPDPDQDYDYSDYEDYNOENTSSLT 60  
QY 61 HAENPDWYTTEDQADPCQPNPCRHGGDCLVHGSTFTCSCLAPFSGNKKQVQNTCKDNP 120  
DB 61 HAENPDWYTTEDQADPCQPNPCRHGGDCLVHGSTFTCSCLAPFSGNKKQVQNTCKDNP 120  
QY 121 GRGQCLITQSPFYRCVCKHPYTGPCSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180  
DB 121 GRGQCLITQSPFYRCVCKHPYTGPCSCQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180  
QY 181 QFKGKFCIEIGDDCYVGDGYSYRGKWRVTNQHACLYNWSHLLLOENYMFEDAETHGI 240  
DB 181 QFKGKFCIEIGDDCYVGDGYSYRGKWRVTNQHACLYNWSHLLLOENYMFEDAETHGI 240  
QY 241 GEHNFRCNPDADEKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPESPTSTKLPGPS 300

DB 241 GEHNFRCNPDADKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300  
QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPQASQSSLPITISMPQGHFCGGALIHPCWVLT 360  
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPQASQSSLPITISMPQGHFCGGALIHPCWVLT 360  
QY 361 AHCTDIKTRHLKVLVDQDLKKEEFHEQSFVQKIFKYSHYNERDEIPHNDIALLLKPV 420  
DB 361 AHCTDIKTRHLKVLVDQDLKKEEFHEQSFVQKIFKYSHYNERDEIPHNDIALLLKPV 420  
QY 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWGVTETGKSRQLLDKVKLIANTLCNSR 480  
DB 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWGVTETGKSRQLLDKVKLIANTLCNSR 480  
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYGVISWGLECKRPGVY 540  
DB 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYGVISWGLECKRPGVY 540  
QY 541 TQVTKFLNWKATIKSESGF 560  
DB 541 TQVTKFLNWKATIKSESGF 560  
RESULT 3  
AAO17144  
ID AAO17144 standard; protein; 560 AA.  
XX AC AAO17144;  
XX DT 13-JUN-2002 (first entry)  
XX DE Human blood coagulation factor VII activating protease.  
XX KW Human; blood coagulation factor VII activating protease; FSAP;  
KW single-chain plasminogen activator; bleeding disorder; haematological;  
KW haemostatic.  
XX OS Homo sapiens.  
XX PN EP1182258-A1.  
XX PD 27-FEB-2002.  
XX PF 05-JUL-2001; 2001EP-00115691.  
XX PR 26-JUL-2000; 2000DE-01036641.  
XX PR 10-OCT-2000; 2000DE-01050040.  
XX PR 21-OCT-2000; 2000DE-01052319.  
XX PR 12-APR-2001; 2001DE-01018706.  
XX PA (AVET ) AVENTIS BEHRING GMBH.  
XX PI Roemisch J, Stoehr H, Feussner A, Lang W, Weimer T, Becker M;  
PI Nerlich C, Muth-Naumann G;  
XX DR WPI; 2002-270939/32.  
XX DR N-PSDB; AAL45696.  
XX PT New nucleic acid encoding mutant factor 7 activating protease, useful for  
PT diagnosis, treatment and prevention of coagulation disorders, also  
PT related protein and antibodies.  
XX PS Claim 4; Page 17-19; 27pp; German.  
XX CC The present invention relates to a mutant of the DNA sequence encoding  
CC the protease (FSAP) that activates blood coagulation factor VII (FVII)  
CC and single-chain plasminogen activator, where at least one of the base  
CC changes G to C at nucleotide 1177 and G to A at nucleotide 1601 is  
CC present. The mutant sequences can be used in the treatment and prevention  
CC of bleeding disorders associated with inherited or acquired defects of  
CC blood coagulation factors V, VIII, IX, X, XI, XII, von Willebrand's  
CC factor and/or with antibodies against any of these proteins. The present

CC sequence is the human FSAP protein  
XX SQ Sequence 560 AA;  
Query Match 99.7%; Score 3143; DB 5; Length 560;  
Best Local Similarity 99.6%; Pred. No. 3 1e-185;  
Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MFARMSDLHLLMALVGKTACGFSLSLLESLLDPDTPDQDYDYEDYNOBENTSSLT 60  
DB 1 MFARMSDLHLLMALVGKTACGFSLSLLESLLDPDTPDQDYDYEDYNOBENTSSLT 60  
QY 61 HAENPMWYTEBOADPCQNPCEHGGDCLVHGSTTCCLAPFSGNKKQVQNTCKNCP 120  
DB 61 HAENPMWYTEBOADPCQNPCEHGGDCLVHGSTTCCLAPFSGNKKQVQNTCKNCP 120  
QY 121 GRQCCLITQSPPYRCVCKHPYTGPSQVQVPCPNPQNGATCSRHRKSKFTCACPD 180  
DB 121 GRQCCLITQSPPYRCVCKHPYTGPSQVQVPCPNPQNGATCSRHRKSKFTCACPD 180  
QY 181 QFKGKFCETGSDDCYVGDGYSYRGKKNRTVNOHACLYWNSHLLLOENYNNWEDAEATHGI 240  
DB 181 QFKGKFCETGSDDCYVGDGYSYRGKKNRTVNOHACLYWNSHLLLOENYNNWEDAEATHGI 240  
QY 241 GEHNFRCNPDADKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300  
DB 241 GEHNFRCNPDADKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEESPTSTKLPGFDS 300  
QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPQASQSSLPITISMPQGHFCGGALIHPCWVLT 360  
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPQASQSSLPITISMPQGHFCGGALIHPCWVLT 360  
QY 361 AHCTDIKTRHLKVLVDQDLKKEEFHEQSFVQKIFKYSHYNERDEIPHNDIALLLKPV 420  
DB 361 AHCTDIKTRHLKVLVDQDLKKEEFHEQSFVQKIFKYSHYNERDEIPHNDIALLLKPV 420  
QY 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWGVTETGKSRQLLDKVKLIANTLCNSR 480  
DB 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWGVTETGKSRQLLDKVKLIANTLCNSR 480  
QY 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYGVISWGLECKRPGVY 540  
DB 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYGVISWGLECKRPGVY 540  
QY 541 TQVTKFLNWKATIKSESGF 560  
DB 541 TQVTKFLNWKATIKSESGF 560  
RESULT 4  
AAR53962  
ID AAR53962 standard; protein; 655 AA.  
XX AC AAR53962;  
XX DT 25-MAR-2003 (revised)  
DT 06-JAN-1995 (first entry)  
XX DE Hepatocyte growth factor converting protease.  
XX KW Hepatocyte growth factor; protease; cleavage; active; inactive;  
KW precursor.  
XX OS Homo sapiens.  
XX PN EP596524-A2.  
XX PD 11-MAY-1994.  
XX PF 05-NOV-1993; 93EP-00117988.  
XX PR 05-NOV-1992; 92JP-00296133.  
PR 20-NOV-1992; 92JP-00312234.

```

PR 20-NOV-1992; 92JP-00312242.
PA (MITU ) MITSUBISHI KASEI CORP.
XX Shimomura T, Yamada K, Morimoto Y, Kitamura N, Miyazawa K;
XX WPT; 1994-152921/19.
XX N-PSDB; AAQ63951.
XX Hepatocyte growth factor converting protease and precursor and gene
XX encoding them - for producing active two chain HGF from inactive single
XX chain HGF.
XX Claim 12; Page 21-24; 30pp; English.
XX Hepatocyte growth factor converting protease is capable of converting
XX inactive single chain hepatocyte growth factor (HGF) into active two
XX chain HGF by cleavage at a specific site. (Updated on 25-MAR-2003 to
XX correct PN field.)
XX Sequence 655 AA;
SQ
Query Match 26.1%; Score 822.5; DB 2; Length 655;
Best Local Similarity 33.2%; Pred. No. 1.6e-42;
Matches 190; Conservative 90; Mismatches 221; Indels 71; Gaps 21;
QY 31 ESLDPDPTDQDYSDYEDYNEENTSSLTTHAENPDW-----YTEDQA----- 74
DB 99 QALTEGRCPPFPFRYGRMLHACTSEGSAHRK---WCATHNYDRDRAWGVCVEATPPP 155
QY 75 -----DPCQNPCEHGGDC--LVHGSTFTCCLAPFSGNKC--QKVONTCK-----D 117
DB 156 GGPALDPCASGCLNGSCSNTQDPQSVHCSCPAFTGKDCGTEKCFDTRVEYLEGGD 215
QY 118 NPCRGQLITQSPYVCVKCHPYTGPSCSQV--PVCPRNCPQCATCSRHKRSKFT 175
DB 216 RWARVRQCHVEQ-----CEC---FGGRTWCETRHTACLSPPCLNGTCHLIVAGTTV 266
QY 176 CACPDQFCKGKEIGSD--CYVGDGYSYRGKMRITVNHQACLWNSHLLQLQENYMPMED 234
DB 267 CACPPGAPAGRLCNIEPDERCFLNGTGYRGVASTSASGLSCLAWNSDLLYOELHVDVSGA 326
QY 235 AETHIGIBHNCRPDADDEKWCICIKVTNDKVKWEYCDVSNCSA--QDVAYPEESPERS 292
DB 327 AALLGLGPHAYCRPNPDERPWCYV-VKDSALSWEYCRLEACEBSLTVQLSPDLLATPE 385
QY 293 TKLPDFDCSGKTEIAERKIK-RIVGGFKSTAGKHPWQASLQSLPLATISMPQGHFCGAL 351
DB 386 PASPORQACGRHKKRTFLRRIIGSSSLPGSHPLAAIYG-----DSFCAGSL 436
QY 352 IHPQVLTAAHCTDIKTRH--LKVLGDQDLKKEEFHQSFVQKIFKYSHYNERDEIPH 409
DB 437 VHTCWVSAAHCFSHSPRDSVSVVLGQHFFNRITTDVTQTGIEKYIPTLYSVFNPSDH 496
QY 410 NDIALLLKVPDGHGCHALSKVKVTCLPD--GSPFSGSECHISGWG-VTEGKG-SROLL 465
DB 497 -DLVIRLKKGDRCATSQFVQICLPEPSTFPAGHKCQIAGWHLDENVSYSLSR 555
QY 466 DARVKLIANTLNSQLYDHMDSDMI CAGNLQKPGQDTCGDSGGPLTCKDGTYYVYG 525
DB 556 EALVPLVADHKCSSPEVYGADISPNMLCAGYFDCK-SDACQDSGGPLACNGVAYLYG 614
QY 526 IYWSGLECEK--RPGVYTVQTKFLANKIKATIK 555
DB 615 IISWGDGCGRLHKPGVYTRVANYVDINDIR 646
RESULT 5
AAR89197
ID AAR89197 standard; protein; 655 AA.
XX
AC
XX AAR89197;
XX

```

```

DT 08-AUG-1996 (first entry)
XX Human hepatocellular growth factor single chain precursor protein.
DE
XX Mature protein; inactive; single chain; hepatocellular growth factor;
KW liver; human; precursor; dimerisation; double chain; serine protease;
XX hepatitis.
XX Homo sapiens.
XX Key Location/Qualifiers
XX FH 356..655
XX Peptide /note="mature peptide"
XX
XX JP08027026-A.
XX
XX 30-JAN-1996.
XX
XX 22-JUL-1994; 94JP-00171207.
XX
XX 22-JUL-1994; 94JP-00171207.
XX
XX (MITU ) MITSUBISHI CHEM CORP.
XX WPI; 1996-136206/14.
XX
XX New preventative and therapeutic cpds contg. a 34 kD serine protease -
XX useful for treatment of liver diseases e.g. hepatitis.
XX Claim 4; Page 6-8; 8pp; Japanese.
XX This is the amino acid sequence of the precursor protein of the inactive
XX single chain form of a hepatocellular growth factor. The mature protein
XX (AAR89196) has a mol. wt. of around 34 kD and is derived from the 96 kD
XX precursor protein. The mature protein corresp. to residues 356-655 of the
XX precursor protein. The inactive single chain form of the growth factor is
XX activated by dimerisation of the mature protein. The active protein is a
XX serine protease which can be used for the treatment of hepatitis
XX
XX Sequence 655 AA;
Query Match 26.1%; Score 821.5; DB 2; Length 655;
Best Local Similarity 32.9%; Pred. No. 1.9e-42;
Matches 190; Conservative 88; Mismatches 217; Indels 83; Gaps 21;
QY 31 ESLDPDPTDQDYSDYEDYNEENTSSLTTHAENPDW-----YTEDQA----- 74
DB 99 QALTEGRCPPFPFRYGRMLHACTSEGSAHRK---WCATHNYDRDRAWGVCVEATPPP 155
QY 75 -----DPCQNPCEHGGDC--LVHGSTFTCCLAPFSGNKCQKVONTCKDNPCGRGQL 126
DB 156 GGPALDPCASGCLNGSCSNTQDPQSVHCSCPAFTG-----XD--CGTEKCF 203
QY 127 ITQSPYY-----RVCXHPYTGSCSQV--PVCPRNCPQCATCSRHK 169
DB 204 DETRYEYLEGGDRWARVRQCHVEQEC---FGGRTWCETRHTACLSPPCLNGTCHLIV 260
QY 170 RRSKFTACPDQFKGFCBIGSD--CYVGDGYSYRGKMRITVNHQACLWNSHLLQLQENY 228
DB 261 AGTTVCACPGFAGRLCNIEPDERCFLNGTGYRGVASTSASGLSCLAWNSDLLYQSLH 320
QY 229 NMFMEADETHGIGEHNFCSNPDADEKPMCFIKVTNDKVKWEYCDVSNCSA--QDVAYPEE 286
DB 321 VDSVGAALLGLGPHAYCRPNPDERPWCYV-VKDSALSWEYCRLEACEBSLTVQLSPDL 379
QY 287 SPTBSTKLPGFDSCKGKTEIAERKIK-RIVGGFKSTAGKHPWQASLQSLPLATISMPQGH 345
DB 380 LATLPEPAGPGRQACGRHKKRTFLRRIIGSSSLPGSHPLAAIYG-----DS 430
QY 346 FCGGALIHFCWLTAAHCTDIKTRH--LKVLGDQDLKKEEFHQSFVQKIFKYSHYNE 403
DB 431 FCAGSLVHTCWVSAAHCFSHSPRDSVSVVLGQHFFNRITTDVTQTGIEKYIPTLYSV 490

```



QY 124 -QCLITQSPYRCVCKHPTGSCQVVPVCEPNPCQ-----NGATCSHKRSKFTCAC 178  
Db 111 AECTNNWS-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDSDSKPWCYV 157  
QY 179 PDQFK-GKF-----CBIGSDDCYVGDYSGYRGKMNRTVNOHACLYWNSHLLLOENY 228  
Db 158 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAVGRTHSLTSGASCLPWNMSMILIGKVY 214  
QY 229 NMFMEDAETHGIGHNFCRNPDADEKFWCFIKVTNDKVKWEYCDVSAGSAQVAYPEESP 288  
Db 215 TAQNPAAQALGLGKHNYCRNPDGDAKPWCHV-LKNRLTWECVDFPSCS-----262  
QY 289 TEPSTKLPQDSCGKTEIARERKIRYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCG 348  
Db 263 -----TCGLROYSQF-RIKGLFADIASHPWQAAIFAK--HRSPPGERFLCG 308  
QY 349 CALIHPWCWLTAHCTD--IKTRHLKVLGDQDLKKEEFHQSFQVQKIFYSHYNERDE 406  
Db 309 GILISSCWILSAARCFQERPPHLLTVILGRTRYVVPGEERQKFEVEKYIVHKEFD--DD 366  
QY 407 IPHNDIALKLKPVGDHGALESKYVTVCLP--DGSFPGSGECHISGMVGTETGK--SR 462  
Db 367 TYNDIALQLKSDSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGKHEALSPFYSE 426  
QY 463 QLLDAKYKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----ODTCQDGGSGPLTCEK 517  
Db 427 RLKENHVRLYPSSRCTSOHLNRTVTDNMLCAGDTRSGGPOANLHDACQDGGSGPLVCLN 486  
QY 518 DGTYYVYGVISWGLECEK--PGVYTVTKPLNMIKATIX 555  
Db 487 DGRMTLVGLIISWGLGCGQKXDPGVYTKVNYLDWIRDNR 526

RESULT 14  
US-09-969-271-7  
; Sequence 7, Application US/09969271  
; Patent No. US20020098179A1  
; GENERAL INFORMATION:  
; APPLICANT: Pfizer Inc. (All designated States except GB and EP (GB));  
; APPLICANT: Pfizer Limited (GB and EP (GB) only)  
; TITLE OF INVENTION: Pharmaceutical Combinations  
; FILE REFERENCE: PCS10951APME  
; CURRENT APPLICATION NUMBER: US/09/969,271  
; PRIOR FILING DATE: 2001-10-01  
; PRIOR APPLICATION NUMBER: GB 0025473.0  
; PRIOR FILING DATE: 2000-10-17  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-969-271-7

Query Match 24.0%; Score 755.5; DB 9; Length 562;  
Best Local Similarity 34.6%; Pred. No. 5.9e-51;  
Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG-----123  
Db 86 CSEPRCFNGGTCCQALYFSDVFCQCEGAGKCCEDTRATCYEDQGISYRGTWSTABSG 145

QY 124 -QCLITQSPYRCVCKHPTGSCQVVPVCEPNPCQ-----NGATCSHKRSKFTCAC 178  
Db 146 AECTNNWS-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDSDSKPWCYV 192

QY 179 PDQFK-GKF-----CBIGSDDCYVGDYSGYRGKMNRTVNOHACLYWNSHLLLOENY 228  
Db 193 ---FKAGKYSSEFCSTPACSEGNSDCYFGNGSAVGRTHSLTSGASCLPWNMSMILIGKVY 249

QY 229 NMFMEDAETHGIGHNFCRNPDADEKFWCFIKVTNDKVKWEYCDVSAGSAQVAYPEESP 288  
Db 250 TAQNPAAQALGLGKHNYCRNPDGDAKPWCHV-LKNRLTWECVDFPSCS-----297

QY 289 TEPSTKLPQDSCGKTEIARERKIRYGGFKSTAGKHPWQASLQSSLPITISMPQGHFCG 348  
Db 298 -----TCGLROYSQF-RIKGLFADIASHPWQAAIFAK--HRSPPGERFLCG 343  
QY 349 GALIHPWCWLTAHCTD--IKTRHLKVLGDQDLKKEEFHQSFQVQKIFYSHYNERDE 406  
Db 344 GILISSCWILSAARCFQERPPHLLTVILGRTRYVVPGEERQKFEVEKYIVHKEFD--DD 401  
QY 407 IPHNDIALKLKPVGDHGALESKYVTVCLP--DGSFPGSGECHISGMVGTETGK--SR 462  
Db 402 TYNDIALQLKSDSRCAQESSVVRTVCLPPADLQLPDWTECELSGYGKHEALSPFYSE 461  
QY 463 QLLDAKYKLIANTLNSRQLYDHMIDDSMICAGNLQKPG-----ODTCQDGGSGPLTCEK 517  
Db 462 RLKENHVRLYPSSRCTSOHLNRTVTDNMLCAGDTRSGGPOANLHDACQDGGSGPLVCLN 521  
QY 518 DGTYYVYGVISWGLECEK--PGVYTVTKPLNMIKATIX 555  
Db 522 DGRMTLVGLIISWGLGCGQKXDPGVYTKVNYLDWIRDNR 561

RESULT 15  
US-10-411-037-26  
; Sequence 26, Application US/10411037  
; Publication No. US20040043446A1  
; GENERAL INFORMATION:  
; APPLICANT: Neose Technologies, Inc.  
; APPLICANT: Deftrees, Shawn  
; APPLICANT: Zopf, David  
; APPLICANT: Beyer, Robert  
; APPLICANT: Hakes, David  
; APPLICANT: Chen, Xi  
; APPLICANT: Bowe, Caryn  
; TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF AL.  
; TITLE OF INVENTION: GALACTOSIDASE A  
; FILE REFERENCE: 040853-01-5082  
; CURRENT APPLICATION NUMBER: US/10/411,037  
; CURRENT FILING DATE: 2003-04-09  
; PRIOR APPLICATION NUMBER: US 60/328,523  
; PRIOR FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 60/344,692  
; PRIOR FILING DATE: 2001-10-19  
; PRIOR APPLICATION NUMBER: US 60/387,292  
; PRIOR FILING DATE: 2002-06-07  
; PRIOR APPLICATION NUMBER: US 60/391,777  
; PRIOR FILING DATE: 2002-06-25  
; PRIOR APPLICATION NUMBER: US 60/396,594  
; PRIOR FILING DATE: 2002-07-17  
; PRIOR APPLICATION NUMBER: US 60/404,249  
; PRIOR FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: US 60/407,527  
; PRIOR FILING DATE: 2002-08-28  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 26  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-411-037-26

Query Match 24.0%; Score 755.5; DB 12; Length 562;  
Best Local Similarity 34.6%; Pred. No. 5.9e-51;  
Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;

QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKKQ-KVQNTCKDNP--CGRG-----123  
Db 86 CSEPRCFNGGTCCQALYFSDVFCQCEGAGKCCEDTRATCYEDQGISYRGTWSTABSG 145

QY 124 -QCLITQSPYRCVCKHPTGSCQVVPVCEPNPCQ-----NGATCSHKRSKFTCAC 178  
Db 146 AECTNNWS-----SALAOKPYSGR-----RPDAIRLGLGNHNYCRNPDSDSKPWCYV 192

Thu May 27 10:53:59 2004

us-09-912-559-4.rapb

NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 19  
LENGTH: 527  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-987-455-19

Query Match 24.0%; Score 755.5; DB 10; Length 527;  
Best Local Similarity 34.6%; Pred. No. 5.4e-51;  
Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;  
Type: PRT  
ORGANISM: Homo sapiens  
US-09-987-455-19

77 QCPNFCHEGGDC--LVHGSTFTCCCLAPPSGKQ--KVQNTCKDNP--CGRG-----123  
51 CSBPCFNGGTCQOQALYFSDVCOCPGEGFAGKCEIDTRATCYEDQGISYRGTWSTAESG 110  
124 -QCLTQSPFYRCVCKHYTGSPSCQVVPVCRNFCQ---NGATCSRHKRSKFTCAC 178  
111 AECTNWS-----SALAQKPSGR-----RPAIRLGLGNHNYCRNDRDPSKPCV 157  
179 PDQPK-GKF-----CEIGSDDCYVDGYSYRGKWNRTVQHACLWNSHLLQENY 228  
158 ---FKAGYSSEFCSTPACSEGNDCYFNGSAYRGTHSLTESGASCLPWSMILIGKY 214  
229 NFMEDAEITGIGHNFCRNPDADKPCFIKVTNDKVKWEYCDVSACSADVAYPEEP 288  
215 TAQNPQAALGLGNHNYCRNDRDPSKPCV---LKNRLTWECYDVPSCS-----262  
289 TEPSTKLPDFDSCGKTEIAERKIRIYGFKSTAGKHPQASLQSSLPITISMPOGHFCG 348  
263 -----TCGLRQYSQPF-RINGGLFADIASHPWQAIAFAK--HRSFGRFLCG 308  
349 GALIHPCWVLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSFVQKIFKYSHYNERDE 406  
309 GILISSCWILSAHCFQERPPHLLTVILGRYRVVPEEEOKFVEKIVHKEFD--DD 366  
407 IPHNDIALKLPVDGHCALSKYKVTCLP--DGSFSGSECHISGKMGVTTGKG--SR 462  
367 TYNDIALQLKSDSRCAQESSVVRVTVCLPPADLQDLPDTECELSYKGHEALSPFYSE 426  
463 QLLDAKVLKLIANTLNSRQLYDHMDDSMICAGNLQKPG-----QDTQGDGGGLTCEK 517  
427 RLKEAHVRLYPSRSTSOHLNRTVTDNMLCAGDTRSGGPOANLHDAQCQDGGGLVCLN 486  
518 DGTYYVYGVISWGLCEK--PGVYVTVTKFLANIKATIK 555  
487 DGRMTLVGIISWGLCGCKDVPVGYTKVTNYLDWIRDNR 526

RESULT 12

US-10-432-842-1  
Sequence 1, Application US/10432842  
Publication No. US20040071707A1  
GENERAL INFORMATION:  
APPLICANT: Veronica A. CARROLL  
APPLICANT: Adrian L. HARRIS  
APPLICANT: Roy BICKNELL  
APPLICANT: Pat PRICE  
TITLE OF INVENTION: MODULATION OF CELL GROWTH  
FILE REFERENCE: 117-450 / N.79507A SER  
CURRENT APPLICATION NUMBER: US/10/432,842  
PRIOR FILING DATE: 2003-09-27  
PRIOR APPLICATION NUMBER: PCT/GB01/05244  
PRIOR FILING DATE: 2001-11-28  
PRIOR APPLICATION NUMBER: GB 0029001.5  
PRIOR FILING DATE: 2000-11-28  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: MS Word  
SEQ ID NO 1  
LENGTH: 527  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-432-842-1

Query Match 24.0%; Score 755.5; DB 12; Length 527;  
Best Local Similarity 34.6%; Pred. No. 5.4e-51;  
Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;  
Type: PRT  
ORGANISM: Homo sapiens  
US-09-987-455-19

77 QCPNFCHEGGDC--LVHGSTFTCCCLAPPSGKQ--KVQNTCKDNP--CGRG-----123  
51 CSBPCFNGGTCQOQALYFSDVCOCPGEGFAGKCEIDTRATCYEDQGISYRGTWSTAESG 110  
124 -QCLTQSPFYRCVCKHYTGSPSCQVVPVCRNFCQ---NGATCSRHKRSKFTCAC 178  
111 AECTNWS-----SALAQKPSGR-----RPAIRLGLGNHNYCRNDRDPSKPCV 157  
179 PDQPK-GKF-----CEIGSDDCYVDGYSYRGKWNRTVQHACLWNSHLLQENY 228  
158 ---FKAGYSSEFCSTPACSEGNDCYFNGSAYRGTHSLTESGASCLPWSMILIGKY 214  
229 NFMEDAEITGIGHNFCRNPDADKPCFIKVTNDKVKWEYCDVSACSADVAYPEEP 288  
215 TAQNPQAALGLGNHNYCRNDRDPSKPCV---LKNRLTWECYDVPSCS-----262  
289 TEPSTKLPDFDSCGKTEIAERKIRIYGFKSTAGKHPQASLQSSLPITISMPOGHFCG 348  
263 -----TCGLRQYSQPF-RINGGLFADIASHPWQAIAFAK--HRSFGRFLCG 308  
349 GALIHPCWVLTAAHCTD--IKTRHLKVLGDQDLKKEEFHQSFVQKIFKYSHYNERDE 406  
309 GILISSCWILSAHCFQERPPHLLTVILGRYRVVPEEEOKFVEKIVHKEFD--DD 366  
407 IPHNDIALKLPVDGHCALSKYKVTCLP--DGSFSGSECHISGKMGVTTGKG--SR 462  
367 TYNDIALQLKSDSRCAQESSVVRVTVCLPPADLQDLPDTECELSYKGHEALSPFYSE 426  
463 QLLDAKVLKLIANTLNSRQLYDHMDDSMICAGNLQKPG-----QDTQGDGGGLTCEK 517  
427 RLKEAHVRLYPSRSTSOHLNRTVTDNMLCAGDTRSGGPOANLHDAQCQDGGGLVCLN 486  
518 DGTYYVYGVISWGLCEK--PGVYVTVTKFLANIKATIK 555  
487 DGRMTLVGIISWGLCGCKDVPVGYTKVTNYLDWIRDNR 526

RESULT 13

US-10-360-101-203  
Sequence 203, Application US/10360101  
Publication No. US20040009550A1  
GENERAL INFORMATION:  
APPLICANT: Moll, Gert N.  
APPLICANT: Leenhouts, Cornelis J.  
TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic  
FILE REFERENCE: 2183-5673  
CURRENT APPLICATION NUMBER: US/10/360,101  
CURRENT FILING DATE: 2003-02-07  
PRIOR APPLICATION NUMBER: EP 02077060.8  
PRIOR FILING DATE: 2002-05-24  
NUMBER OF SEQ ID NOS: 309  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 203  
LENGTH: 527  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: sequence of alteplase  
US-10-360-101-203

Query Match 24.0%; Score 755.5; DB 15; Length 527;  
Best Local Similarity 34.6%; Pred. No. 5.4e-51;  
Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;  
Type: PRT  
ORGANISM: Homo sapiens  
US-10-432-842-1

77 QCPNFCHEGGDC--LVHGSTFTCCCLAPPSGKQ--KVQNTCKDNP--CGRG-----123  
51 CSBPCFNGGTCQOQALYFSDVCOCPGEGFAGKCEIDTRATCYEDQGISYRGTWSTAESG 110

GENERAL INFORMATION:  
; APPLICANT: Chen, Hwei-Mei  
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER  
; FILE REFERENCE: PA-0037 P  
; CURRENT APPLICATION NUMBER: US/09/974,298  
; CURRENT FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: 60/238,331  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: PERL Program  
; SEQ ID NO 145  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20020156263A1 1001470CD1  
US-09-974-298-145

Query Match 24.0%; Score 756.5; DB 9; Length 562;  
Best Local Similarity 34.6%; Pred. No. 4.9e-51;  
Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;  
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKCO-KVQNTCKDNP--CGRG-----123  
Db 86 CSEPRCFNGGTCQQAALYFSDVFCQCEGAGKCCCEIDTRATCYBQGGISYRGTWSTAESG 145  
QY 124 -QCLITQSPPYRCVCKHPYTGPCSQVVPVCRPNPCQ----NGATCSRHRKRSKFTCAC 178  
Db 146 AECTNWS---SALAQKPYSGR-----RFDALRLGLGNHNYCRNPDRDSKPCVY 192  
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMTVNOHACLWNSHLLQENY 228  
Db 193 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAYRGTHSLTSGASCLWNSMILIGKY 249  
QY 229 NFMWEDAETHGIGHNFCRNPDADEKPCWCFIKVTDNKWECYDVSAQADVAYPEESP 288  
Db 250 TAQNPQAALGLGKHNYCRNPDCGAKPCHV-LKNERLTWEYCDVPSCS-----297  
QY 289 TEPSTKLPGDSCGKTEIAERKIKRYGGFKSTAGHPQOASLQSSLPITISMPQHFPG 348  
Db 298 -----TCGLRQYSQPF-RINGGLFADIASHPWQAIAFAK--HRRSPGERFLCG 343  
QY 349 GALIHPCWLTAAHCTD--IKTRHLKVLGDQDLKKEEFHEQSFYVKIKFYSHYNERDE 406  
Db 344 GILISSCWILSAACHCFQERPPPHLTVILGRTYRVVPGEEQKFEVEKIVHKEFD--DD 401  
QY 407 IPHNDIALLLKVPDGHCALESKYVTVCLP--DGSFPGSECHIISGWVTETGKG--SR 462  
Db 402 TYNDIALLLQKSDSRCAQESSVVRTVCLPPADLQLPDWTCELSGYGKHEALSFPYSE 461  
QY 463 QLLDAKVKLIANTLCSRQLYDHMDSDMTCAGNLQKPG-----QDTCQDSSGGLTCEK 517  
Db 462 RLKEAHLVLPSSRCTSHLLNRTVTDNMLCAGDTRSGGPOANLHDACQDSSGGLVCLN 521  
QY 518 DGTYYVYGVISWGLECEKR--PGVYTVQTKFLNWKATIK 555  
Db 522 DGRMTLVGIISWGLCGCKQKDPGVYTVKTVYLDWIRDNR 561

RESULT 10  
US-09-987-457-18  
; Sequence 18, Application US/09987457  
; Publication No. US20030013150A1  
; GENERAL INFORMATION:  
; APPLICANT: Manosroi, Aranya  
; APPLICANT: Manosroi, Jiradej  
; APPLICANT: Tayapiwatana, Chatchai  
; APPLICANT: Goetz, Friedrich  
; APPLICANT: Weirer, Rolf-Guenther  
; TITLE OF INVENTION: Methods for Large Scale Protein Production in Prokaryotes  
; FILE REFERENCE: 0652.218001  
; CURRENT APPLICATION NUMBER: US/09/987,457

CURRENT FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/268,573  
; PRIOR FILING DATE: 2001-02-15  
; PRIOR APPLICATION NUMBER: GB 00 27 782.2  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 18  
; LENGTH: 527  
; TYPE: PRT  
; ORGANISM: Homo sapiens (tPA)  
US-09-987-457-18  
Query Match 24.0%; Score 755.5; DB 10; Length 527;  
Best Local Similarity 34.6%; Pred. No. 5.4e-51;  
Matches 180; Conservative 78; Mismatches 177; Indels 85; Gaps 20;  
QY 77 CQNPCEHGGDC--LVHGSTFTCSCLAPFSGNKCO-KVQNTCKDNP--CGRG-----123  
Db 51 CSEPRCFNGGTCQQAALYFSDVFCQCEGAGKCCCEIDTRATCYEDQGISYRGTWSTAESG 110  
QY 124 -QCLITQSPPYRCVCKHPYTGPCSQVVPVCRPNPCQ----NGATCSRHRKRSKFTCAC 178  
Db 111 AECTNWS---SALAQKPYSGR-----RFDALRLGLGNHNYCRNPDRDSKPCVY 157  
QY 179 PDQFK-GKF-----CEIGSDDCYVGDGYSYRGKMTVNOHACLWNSHLLQENY 228  
Db 158 ---FKAGKYSSEFCSTPACSEGNDCYFGNGSAYRGTHSLTSGASCLWNSMILIGKY 214  
QY 229 NFMWEDAETHGIGHNFCRNPDADEKPCWCFIKVTDNKWECYDVSAQADVAYPEESP 288  
Db 215 TAQNPQAALGLGKHNYCRNPDCGAKPCHV-LKNERLTWEYCDVPSCS-----262  
QY 289 TEPSTKLPGDSCGKTEIAERKIKRYGGFKSTAGHPQOASLQSSLPITISMPQHFPG 348  
Db 263 -----TCGLRQYSQPF-RINGGLFADIASHPWQAIAFAK--HRRSPGERFLCG 308  
QY 349 GALIHPCWLTAAHCTD--IKTRHLKVLGDQDLKKEEFHEQSFYVKIKFYSHYNERDE 406  
Db 309 GILISSCWILSAACHCFQERPPPHLTVILGRTYRVVPGEEQKFEVEKIVHKEFD--DD 366  
QY 407 IPHNDIALLLKVPDGHCALESKYVTVCLP--DGSFPGSECHIISGWVTETGKG--SR 462  
Db 367 TYNDIALLLQKSDSRCAQESSVVRTVCLPPADLQLPDWTCELSGYGKHEALSFPYSE 426  
QY 463 QLLDAKVKLIANTLCSRQLYDHMDSDMTCAGNLQKPG-----QDTCQDSSGGLTCEK 517  
Db 427 RLKEAHLVLPSSRCTSHLLNRTVTDNMLCAGDTRSGGPOANLHDACQDSSGGLVCLN 486  
QY 518 DGTYYVYGVISWGLECEKR--PGVYTVQTKFLNWKATIK 555  
Db 487 DGRMTLVGIISWGLCGCKQKDPGVYTVKTVYLDWIRDNR 526

RESULT 11  
US-09-987-455-19  
; Sequence 19, Application US/09987455  
; Publication No. US20030049729A1  
; GENERAL INFORMATION:  
; APPLICANT: Aranya, Manosroi  
; APPLICANT: Jiradej, Manosroi  
; APPLICANT: Chatchai, Tayapiwatana  
; APPLICANT: Friedrich, Goetz  
; APPLICANT: Rolf-Guenther, Werner  
; TITLE OF INVENTION: Methods for Large Scale Production of Recombinant  
; TITLE OF INVENTION: DNA-Derived tPA or K2S Molecules  
; FILE REFERENCE: 0652.219001  
; CURRENT APPLICATION NUMBER: US/09/987,455  
; CURRENT FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/268,574  
; PRIOR FILING DATE: 2001-02-15  
; PRIOR APPLICATION NUMBER: GB 0027779.8  
; PRIOR FILING DATE: 2000-11-14

; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF  
 ; APPLICANT: ROEMISCH, JUERGEN  
 ; APPLICANT: WEIMER, THOMAS  
 ; APPLICANT: FEUSSNER, ANNETTE  
 ; APPLICANT: STOHR, HANS-ARNOLD  
 ; APPLICANT: DOERSAM, VOLKER  
 ; TITLE OF INVENTION: MABURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE  
 ; FILE REFERENCE: (PSP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS  
 ; CURRENT APPLICATION NUMBER: US/10/391,215  
 ; CURRENT FILING DATE: 2003-03-19  
 ; PRIOR APPLICATION NUMBER: 09/912,559  
 ; PRIOR FILING DATE: 2001-07-26  
 ; PRIOR APPLICATION NUMBER: DE 100 36 641.4  
 ; PRIOR FILING DATE: 2000-07-26  
 ; PRIOR APPLICATION NUMBER: DE 100 50 040.4  
 ; PRIOR FILING DATE: 2000-10-10  
 ; PRIOR APPLICATION NUMBER: DE 100 52 319.6  
 ; PRIOR FILING DATE: 2000-10-21  
 ; PRIOR APPLICATION NUMBER: DE 101 18 706.8  
 ; PRIOR FILING DATE: 2001-04-12  
 ; NUMBER OF SEQ ID NOS: 8  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 560  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-391-215-5

Query Match 99.7%; Score 3143; DB 15; Length 560;  
 Best Local Similarity 99.6%; Pred. No. 4e-239;  
 Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MFARMSDLHLLMALVGTACGFSLSLLESLLSDPDWTDPOYDYSYEDYNQENSSSTLT 60  
 DB 1 MFARMSDLHLLMALVGTACGFSLSLLESLLSDPDWTDPOYDYSYEDYNQENSSSTLT 60  
 QY 61 HAENPDWYTTDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKKQKQVNTCKDNPC 120  
 DB 61 HAENPDWYTTDQADPCQNPCEHGGDCLVHGSTFTCSCLAPFSGNKKQKQVNTCKDNPC 120  
 QY 121 GRGQCLITQSPPPYRCVCKHPYTGPSQSVVPCRNPNCONGATCSRHKRSKFTCACPD 180  
 DB 121 GRGQCLITQSPPPYRCVCKHPYTGPSQSVVPCRNPNCONGATCSRHKRSKFTCACPD 180  
 QY 181 QFKGKCEIGSDCYVGDGYSYRGKMNRTVNHQACLYWNSHLLLOENYNMFVEDAETHGI 240  
 DB 181 QFKGKCEIGSDCYVGDGYSYRGKMNRTVNHQACLYWNSHLLLOENYNMFVEDAETHGI 240  
 QY 241 GEHNFENPDADKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEESPSTKLPQFDS 300  
 DB 241 GEHNFENPDADKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEESPSTKLPQFDS 300  
 QY 301 CGKTEIAERKIKRTYGGFKSTAGHPQWASLQSSLPITISMPQGHFCGGALHPCWVLT 360  
 DB 301 CGKTEIAERKIKRTYGGFKSTAGHPQWASLQSSLPITISMPQGHFCGGALHPCWVLT 360  
 QY 361 AHCTDIKTRHLKVLGQDQKLKBEFHEQSPRVQKIFKYSHYNERDRIPHNDAALLKLPV 420  
 DB 361 AHCTDIKTRHLKVLGQDQKLKBEFHEQSPRVQKIFKYSHYNERDRIPHNDAALLKLPV 420  
 QY 421 DGHCALESKVKVTKVCLPDGSPFSGSECHISGWTETGKSGRQLLDAKVKLIANTLCNSR 480  
 DB 421 DGHCALESKVKVTKVCLPDGSPFSGSECHISGWTETGKSGRQLLDAKVKLIANTLCNSR 480  
 QY 481 QLYDHMTDDSMICAGNLQKPGQDTCQDSDGSGPLTCEKDGTYVYVGVSWGLECEKPGVY 540  
 DB 481 QLYDHMTDDSMICAGNLQKPGQDTCQDSDGSGPLTCEKDGTYVYVGVSWGLECEKPGVY 540  
 QY 541 TQVTKFLNWKATIKSSSGF 560  
 DB 541 TQVTKFLNWKATIKSSSGF 560

RESULT 8  
 US-10-172-712-28  
 ; Sequence 28, Application US/10172712  
 ; Publication No. US20030125232A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GRIFFIN, JOHN H.  
 ; APPLICANT: GALE, ANDREW J.  
 ; APPLICANT: GETZOFF, ELIZABETH D.  
 ; APPLICANT: PELLEQUER, JEAN-LUC  
 ; TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS  
 ; FILE REFERENCE: 4198-4001US1  
 ; CURRENT APPLICATION NUMBER: US/10/172,712  
 ; CURRENT FILING DATE: 2002-09-30  
 ; PRIOR APPLICATION NUMBER: 60/298,578  
 ; PRIOR FILING DATE: 2001-06-14  
 ; NUMBER OF SEQ ID NOS: 32  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 28  
 ; LENGTH: 655  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-172-712-28

Query Match 26.1%; Score 821.5; DB 14; Length 655;  
 Best Local Similarity 32.9%; Pred. No. 4.5e-56;  
 Matches 190; Conservative 88; Mismatches 217; Indels 83; Gaps 21;  
 QY 31 ESIDPDWTPDOYDYSYEDYNQENSSSTLTHAENPDW-----YVTEQDA----- 74  
 DB 99 QALTEGRPCRFPPRYGGRMLHACTSGSAHRK---WCATHYVDRDRANGYCVETATPPP 155  
 QY 75 -----DPCPNPCEHGGDCLVHGSTFTCSCLAPFSGNKKQKQVNTCKDNPCGRQOCL 126  
 DB 156 GGPAAALDPCASGCLNGSGCSNTQDPOSYHSCSPRAFTG-----XD--CGTEKCF 203  
 QY 127 ITQSPPPY-----RCVCKHPYTGPSQSVV--PVCRNPNCONGATCSRHK 169  
 DB 204 DETRYLEGGDRWARVRQGHVEQCEC---FGGTWCEGTRHTACLSPLNGTCHLIV 260  
 QY 170 RRSKFTACAPQDFKGFCEIGSDD-CYVGDGYSYRGKMNRTVNHQACLYWNSHLLLOENY 228  
 DB 261 ATGTTVCACPPGFAGRLCNIEPDERCFLNGTGYRGVASTSASGLSCLAWNSDLLYQELH 320  
 QY 229 NMFVEDAETHGIGHNFCRNPDADKPCWCFIKVTNDKVKWEYCDVSACSADQVAYPEE 286  
 DB 321 VDSVGAALLGLGPHAYCRNPNDNERPWCYV-VKDSALSWEYCRLEACESLTRYQLSPDL 379  
 QY 287 SPTBSTKLPQFSDCGKTEIAERKIK-RIYGGFKSTAGHPWQASLQSSLPITISMPQGH 345  
 DB 380 LATLPEPASPRQACGRHKKTFLRPIIGSSSLFGSHPLAAIYG-----DS 430  
 QY 346 FCGGALHPCWVLTAAHCTDIKTRH--LKVVLGQDQKLKBEFHEQSPRVQKIFKYSHYNE 403  
 DB 431 FCAGSLVHTCWWVSAAHCFSSHPPRDSVSVVLGQHFFNRITTDVQTGIEKYIPTYLSV 490  
 QY 404 RDETPHNDIALKLPVDGHCALSKVKVTKVCLPD--GSPSPSGSECHISGWTETGK 460  
 DB 491 FNPSDH-DLVILRLKKGDRCATRSQFVQICLPEGSTFPAGHKCQIAGHGHLDENVSG 549  
 QY 461 -SRQLDAKVKLIANTLCNSRQLYDHMTDDSMICAGNLQKPGQDTCQDSDGSGPLTCEKDG 519  
 DB 550 YSSSLREALVPLVADHKCSSPEVYGADISPNWLCAGYFDCK-SDACQDSDGSGPLACXNG 608  
 QY 520 TYVYVGVSWGLECEK--RPGVYTVQVTKFLNWKATIK 555  
 DB 609 VAYLYGIISWGGGRLKPGVYTRVANYVWINDRIR 646

RESULT 9  
 US-09-974-298-145  
 ; Sequence 145, Application US/09974298  
 ; Patent No. US20020156263A1

APPLICANT: LANG, WIEGAND  
APPLICANT: WEINER, THOMAS  
APPLICANT: BECKER, MARGRET  
APPLICANT: NERLICH, CLAUDIA  
APPLICANT: MUTH-NAUMANN, GUDRUN  
TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND  
DETECTION METHODS USING SPECIFIC ANTIBODIES  
FILE REFERENCE: 06478.1457  
CURRENT APPLICATION NUMBER: US/09/912.559  
CURRENT FILING DATE: 2001-07-26  
PRIOR APPLICATION NUMBER: DE 100 36 541.4  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: DE 100 50 040.4  
PRIOR FILING DATE: 2000-10-10  
PRIOR APPLICATION NUMBER: DE 100 52 319.6  
PRIOR FILING DATE: 2000-10-21  
PRIOR APPLICATION NUMBER: DE 101 18 706.8  
PRIOR FILING DATE: 2001-04-12  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-912-559-3

Query Match 99.7%; Score 3143; DB 9; Length 560;  
Best Local Similarity 99.6%; Pred. No. 4e-239;  
Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 MFARMSDLHVLMLALVGKTCAGFSLMSLLESLEDPDWTDPQDYDYSEDYDYNQENSTSLT 60  
Db 1 MFARMSDLHVLMLALVGKTCAGFSLMSLLESLEDPDWTDPQDYDYSEDYDYNQENSTSLT 60  
Qy 61 HAENPDWYTTEDQADPCPNCEHGGDCLVHGSTFTCSCLAPFSGNKCQKQVNTCKDNPC 120  
Db 61 HAENPDWYTTEDQADPCPNCEHGGDCLVHGSTFTCSCLAPFSGNKCQKQVNTCKDNPC 120  
Qy 121 GRGQCLITQSPYYRCVCKHPYTGSCQVVPVCPNCPQNGATCSRHKRSKFTACPD 180  
Db 121 GRGQCLITQSPYYRCVCKHPYTGSCQVVPVCPNCPQNGATCSRHKRSKFTACPD 180  
Qy 181 QFKGKFCIGSDDCYVGDGYSGYRGKMNRTVNOHACLWNSHLLLOENYMFMEDEATHGI 240  
Db 181 QFKGKFCIGSDDCYVGDGYSGYRGKMNRTVNOHACLWNSHLLLOENYMFMEDEATHGI 240  
Qy 241 GEHNFERNPDADKPCWCFIKVTNDVKWEYCDVSACSADQVAYPEESTPSTKLPFGDS 300  
Db 241 GEHNFERNPDADKPCWCFIKVTNDVKWEYCDVSACSADQVAYPEESTPSTKLPFGDS 300  
Qy 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSPLTISMPQGHFCGALIHPCWVLT 360  
Db 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSPLTISMPQGHFCGALIHPCWVLT 360  
Qy 361 AHCTDIKTRHLKVVLGDQDLKKEEFHEQSFVQKIFKYSHYNERDEIPHNDIALLKLPV 420  
Db 361 AHCTDIKTRHLKVVLGDQDLKKEEFHEQSFVQKIFKYSHYNERDEIPHNDIALLKLPV 420  
Qy 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWSGTETGKSRQLLDKVKLIANTLCNSR 480  
Db 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWSGTETGKSRQLLDKVKLIANTLCNSR 480  
Qy 481 QLYDHMIDSMICAGNLQKPGQDTQCGDSGGPLTCEKDGTYVYVGVISWGLECEKRPVY 540  
Db 481 QLYDHMIDSMICAGNLQKPGQDTQCGDSGGPLTCEKDGTYVYVGVISWGLECEKRPVY 540  
Qy 541 TQVTKFLNWKATIKSESGF 560  
Db 541 TQVTKFLNWKATIKSESGF 560

RESULT 6  
US-10-172-712-32

Sequence 32, Application US/10172712  
Publication No. US20030125232A1  
GENERAL INFORMATION:  
APPLICANT: GRIFFIN, JOHN H.  
APPLICANT: GALE, ANDREW J.  
APPLICANT: GEIZOFF, ELIZABETH D.  
APPLICANT: PELUEQUER, JEAN-LUC  
TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS  
FILE REFERENCE: 4198-4001US1  
CURRENT APPLICATION NUMBER: US/10/172.712  
CURRENT FILING DATE: 2002-09-30  
PRIOR APPLICATION NUMBER: 60/298,578  
PRIOR FILING DATE: 2001-06-14  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 32  
LENGTH: 560  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-172-712-32

Query Match 99.7%; Score 3143; DB 14; Length 560;  
Best Local Similarity 99.6%; Pred. No. 4e-239;  
Matches 558; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 MFARMSDLHVLMLALVGKTCAGFSLMSLLESLEDPDWTDPQDYDYSEDYDYNQENSTSLT 60  
Db 1 MFARMSDLHVLMLALVGKTCAGFSLMSLLESLEDPDWTDPQDYDYSEDYDYNQENSTSLT 60  
Qy 61 HAENPDWYTTEDQADPCPNCEHGGDCLVHGSTFTCSCLAPFSGNKCQKQVNTCKDNPC 120  
Db 61 HAENPDWYTTEDQADPCPNCEHGGDCLVHGSTFTCSCLAPFSGNKCQKQVNTCKDNPC 120  
Qy 121 GRGQCLITQSPYYRCVCKHPYTGSCQVVPVCPNCPQNGATCSRHKRSKFTACPD 180  
Db 121 GRGQCLITQSPYYRCVCKHPYTGSCQVVPVCPNCPQNGATCSRHKRSKFTACPD 180  
Qy 181 QFKGKFCIGSDDCYVGDGYSGYRGKMNRTVNOHACLWNSHLLLOENYMFMEDEATHGI 240  
Db 181 QFKGKFCIGSDDCYVGDGYSGYRGKMNRTVNOHACLWNSHLLLOENYMFMEDEATHGI 240  
Qy 241 GEHNFERNPDADKPCWCFIKVTNDVKWEYCDVSACSADQVAYPEESTPSTKLPFGDS 300  
Db 241 GEHNFERNPDADKPCWCFIKVTNDVKWEYCDVSACSADQVAYPEESTPSTKLPFGDS 300  
Qy 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSPLTISMPQGHFCGALIHPCWVLT 360  
Db 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLQSSPLTISMPQGHFCGALIHPCWVLT 360  
Qy 361 AHCTDIKTRHLKVVLGDQDLKKEEFHEQSFVQKIFKYSHYNERDEIPHNDIALLKLPV 420  
Db 361 AHCTDIKTRHLKVVLGDQDLKKEEFHEQSFVQKIFKYSHYNERDEIPHNDIALLKLPV 420  
Qy 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWSGTETGKSRQLLDKVKLIANTLCNSR 480  
Db 421 DGHCALESKYVKTVCPLPDGSPFSGSECHISGWSGTETGKSRQLLDKVKLIANTLCNSR 480  
Qy 481 QLYDHMIDSMICAGNLQKPGQDTQCGDSGGPLTCEKDGTYVYVGVISWGLECEKRPVY 540  
Db 481 QLYDHMIDSMICAGNLQKPGQDTQCGDSGGPLTCEKDGTYVYVGVISWGLECEKRPVY 540  
Qy 541 TQVTKFLNWKATIKSESGF 560  
Db 541 TQVTKFLNWKATIKSESGF 560

RESULT 7  
US-10-391-215-5  
Sequence 5, Application US/10391215  
Publication No. US20040009543A1  
GENERAL INFORMATION:  
APPLICANT: KIECHL, STEFAN  
APPLICANT: WILLEIT, JOHANN

; PRIOR APPLICATION NUMBER: DE 100 52 319.6  
; PRIOR FILING DATE: 2000-10-21  
; PRIOR APPLICATION NUMBER: DE 101 18 706.8  
; PRIOR FILING DATE: 2001-04-12  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 560  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-391-215-6

Query Match 99.9%; Score 3150; DB 15; Length 560;  
Best Local Similarity 99.8%; Pred. No. 1.1e-239;  
Matches 559; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLALVGKTACGFSLSLLESLLSDPDWTPDQDYDYSYEDYNOBENTSSLT 60  
DB 1 MFARMSDLHVLMLALVGKTACGFSLSLLESLLSDPDWTPDQDYDYSYEDYNOBENTSSLT 60

QY 61 HAENPDWYTTEDQADPCQPNPCHEHGGDCLVHGSTFTCSCLAPFSGNKCQKVQNTCKDNPC 120  
DB 61 HAENPDWYTTEDQADPCQPNPCHEHGGDCLVHGSTFTCSCLAPFSGNKCQKVQNTCKDNPC 120

QY 121 GRGQCLITQSPPYRCVCKHPYTGSCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180  
DB 121 GRGQCLITQSPPYRCVCKHPYTGSCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

QY 181 QFKGFCEIGSDDCVVGDGYSYRGKNRTVNOHACLWNSHLLLOENYMFMEDEATHGI 240  
DB 181 QFKGFCEIGSDDCVVGDGYSYRGKNRTVNOHACLWNSHLLLOENYMFMEDEATHGI 240

QY 241 GEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEESTPSTKLPGFDS 300  
DB 241 GEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEESTPSTKLPGFDS 300

QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSPLTISMPQGHFCGGLIHPCWVLT 360  
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSPLTISMPQGHFCGGLIHPCWVLT 360

QY 361 AHCTDIKTRHLKVVLGDQDLKKEEFHEQSFVOKIFKYSHYNERDEIPHNDIALLLKXPV 420  
DB 361 AHCTDIKTRHLKVVLGDQDLKKEEFHEQSFVOKIFKYSHYNERDEIPHNDIALLLKXPV 420

QY 421 DGHCALESKYVKTVCPLPDGSPGSECHISGNGVTETGKSRQLLDKVKLIANTLCNSR 480  
DB 421 DGHCALESKYVKTVCPLPDGSPGSECHISGNGVTETGKSRQLLDKVKLIANTLCNSR 480

QY 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVYGVISWGLECEKRPVY 540  
DB 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVYGVISWGLECEKRPVY 540

QY 541 TQVTKFLNWKATIKSES GF 560  
DB 541 TQVTKFLNWKATIKSES GF 560

RESULT 4  
US-10-391-215-7  
; Sequence 7, Application US/10391215  
; Publication No. US20040009543A1  
; GENERAL INFORMATION:  
; APPLICANT: KIECHL, STEFAN  
; APPLICANT: WILLEIT, JOHANN  
; APPLICANT: WIEDERMANN, CHRISTIAN JOSEF  
; APPLICANT: ROEMISCH, JUERGEN  
; APPLICANT: WEIMER, THOMAS  
; APPLICANT: FEUSSNER, ANNETTE  
; APPLICANT: STOEHR, HANS-ARNOLD  
; APPLICANT: DOERSAM, VOLKER  
; TITLE OF INVENTION: MARBURG I MUTANT OF FACTOR VII ACTIVATING PROTEASE  
; TITLE OF INVENTION: (FSAP) AS RISK FACTOR FOR ARTERIAL THROMBOSIS  
; FILE REFERENCE: 06478.1457-01

; CURRENT APPLICATION NUMBER: US/10/391,215  
; CURRENT FILING DATE: 2003-03-19  
; PRIOR APPLICATION NUMBER: 09/912,559  
; PRIOR FILING DATE: 2001-07-26  
; PRIOR APPLICATION NUMBER: DE 100 36 641.4  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: DE 100 50 040.4  
; PRIOR FILING DATE: 2000-10-10  
; PRIOR APPLICATION NUMBER: DE 100 52 319.6  
; PRIOR FILING DATE: 2000-10-21  
; PRIOR APPLICATION NUMBER: DE 101 18 706.8  
; PRIOR FILING DATE: 2001-04-12  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 560  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-391-215-7

Query Match 99.8%; Score 3146; DB 15; Length 560;  
Best Local Similarity 99.8%; Pred. No. 2.3e-239;  
Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFARMSDLHVLMLALVGKTACGFSLSLLESLLSDPDWTPDQDYDYSYEDYNOBENTSSLT 60  
DB 1 MFARMSDLHVLMLALVGKTACGFSLSLLESLLSDPDWTPDQDYDYSYEDYNOBENTSSLT 60

QY 61 HAENPDWYTTEDQADPCQPNPCHEHGGDCLVHGSTFTCSCLAPFSGNKCQKVQNTCKDNPC 120  
DB 61 HAENPDWYTTEDQADPCQPNPCHEHGGDCLVHGSTFTCSCLAPFSGNKCQKVQNTCKDNPC 120

QY 121 GRGQCLITQSPPYRCVCKHPYTGSCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180  
DB 121 GRGQCLITQSPPYRCVCKHPYTGSCSQVVPVCRPNPCQNGATCSRHKRSKFTACPD 180

QY 181 QFKGFCEIGSDDCVVGDGYSYRGKNRTVNOHACLWNSHLLLOENYMFMEDEATHGI 240  
DB 181 QFKGFCEIGSDDCVVGDGYSYRGKNRTVNOHACLWNSHLLLOENYMFMEDEATHGI 240

QY 241 GEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEESTPSTKLPGFDS 300  
DB 241 GEHNFRCNPDADEKPCWCFIKVTNDKVWEYCDVSACSADQVAYPEESTPSTKLPGFDS 300

QY 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSPLTISMPQGHFCGGLIHPCWVLT 360  
DB 301 CGKTEIAERKIKRIYGGFKSTAGKHPWQASLOSPLTISMPQGHFCGGLIHPCWVLT 360

QY 361 AHCTDIKTRHLKVVLGDQDLKKEEFHEQSFVOKIFKYSHYNERDEIPHNDIALLLKXPV 420  
DB 361 AHCTDIKTRHLKVVLGDQDLKKEEFHEQSFVOKIFKYSHYNERDEIPHNDIALLLKXPV 420

QY 421 DGHCALESKYVKTVCPLPDGSPGSECHISGNGVTETGKSRQLLDKVKLIANTLCNSR 480  
DB 421 DGHCALESKYVKTVCPLPDGSPGSECHISGNGVTETGKSRQLLDKVKLIANTLCNSR 480

QY 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVYGVISWGLECEKRPVY 540  
DB 481 QLYDHMIDDSMICAGNLQKPGQDTCCGDSGGPLTCEKDGTYVYVYGVISWGLECEKRPVY 540

QY 541 TQVTKFLNWKATIKSES GF 560  
DB 541 TQVTKFLNWKATIKSES GF 560

RESULT 5  
US-09-912-559-3  
; Sequence 3, Application US/09912559  
; Patent No. US20020142316A1  
; GENERAL INFORMATION:  
; APPLICANT: ROEMISCH, JUERGEN  
; APPLICANT: STOEHR, HANS-ARNOLD  
; APPLICANT: FEUSSNER, ANNETTE





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 24, 2004, 09:49:06 ; Search time 42 Seconds  
(without alignments)  
3718.956 Million cell updates/sec

Title: US-09-912-559-4

Perfect score: 3153

Sequence: 1 MFARMSDLHVLMLALVGKT.....TQVTKFLNWKATIKSBSGF 560

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3153	100.0	560	9	US-09-912-559-4
2	3153	100.0	560	15	US-10-391-215-8
3	3150	99.9	560	15	US-10-391-215-6
4	3146	99.8	560	15	US-10-391-215-7
5	3143	99.7	560	9	US-09-912-559-3
6	3143	99.7	560	14	US-10-172-712-32
7	3143	99.7	560	15	US-10-391-215-5
8	821.5	26.1	655	14	US-10-172-712-28
9	756.5	24.0	562	9	US-09-974-298-145
10	755.5	24.0	527	10	US-09-987-457-18
11	755.5	24.0	527	10	US-09-987-455-19
12	755.5	24.0	527	12	US-10-432-842-1
13	755.5	24.0	527	15	US-10-360-101-203
14	755.5	24.0	562	9	US-09-969-271-7
15	755.5	24.0	562	12	US-10-411-037-26

16	755.5	24.0	562	12	US-10-411-026-26
17	755.5	24.0	562	14	US-10-193-656-8
18	755.5	24.0	562	14	US-10-443-701-4
19	755.5	24.0	562	16	US-10-410-962-26
20	755.5	24.0	562	16	US-10-411-049-26
21	742.5	23.5	650	15	US-10-401-077-1
22	733	23.2	615	10	US-09-858-909-2
23	733	23.2	615	14	US-10-172-712-30
24	733	23.2	615	16	US-10-449-132-2
25	708	22.5	372	9	US-09-084-491A-3
26	708	22.5	372	13	US-10-102-704-3
27	698.5	22.2	354	10	US-09-987-457-10
28	698.5	22.2	354	10	US-09-987-455-11
29	698.5	22.2	377	10	US-09-987-455-8
30	677	21.5	433	12	US-10-087-192-591
31	660.5	20.9	411	15	US-10-407-821-2
32	659.5	20.9	411	9	US-09-880-503-3
33	659.5	20.9	431	12	US-10-411-037-34
34	659.5	20.9	431	12	US-10-411-026-34
35	659.5	20.9	431	13	US-10-076-421-2
36	659.5	20.9	431	14	US-10-171-311-184
37	659.5	20.9	431	14	US-10-301-822-161
38	659.5	20.9	431	14	US-10-131-985-21
39	659.5	20.9	431	15	US-10-295-027-414
40	659.5	20.9	431	15	US-10-295-027-1275
41	659.5	20.9	431	16	US-10-410-962-34
42	659.5	20.9	431	16	US-10-411-049-34
43	659.5	20.9	437	12	US-10-087-192-594
44	659	20.9	812	12	US-09-825-751A-71
45	658.5	20.9	431	14	US-10-247-671-149

ALIGNMENTS

RESULT 1

US-09-912-559-4

Sequence 4, Application US/09912559

Patent No. US20020142316A1

GENERAL INFORMATION:

APPLICANT: ROEMISCH, JUERGEN

APPLICANT: STOEHR, HANS-ARNOLD

APPLICANT: FEUSSNER, ANNETTE

APPLICANT: LANG, WIEGAND

APPLICANT: WEIMER, THOMAS

APPLICANT: BECKER, MARGRET

APPLICANT: NESLICH, CLAUDIA

APPLICANT: MUTH-NAUMANN, GUDRUN

TITLE OF INVENTION: MUTANTS OF THE FACTOR VII-ACTIVATING PROTEASE AND

FILE REFERENCE: 06478.1457

CURRENT APPLICATION NUMBER: US/09/912.559

PRIOR FILING DATE: 2001-07-26

PRIOR FILING DATE: 2000-07-26

PRIOR FILING DATE: 2000-07-26

PRIOR FILING DATE: 2000-07-26

PRIOR FILING DATE: 2000-10-10

PRIOR FILING DATE: 2000-10-21

PRIOR FILING DATE: 2000-10-21

PRIOR FILING DATE: 2001-04-12

NUMBER OF SEQ ID NOS: 4

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 4

LENGTH: 560

TYPE: PRT

ORGANISM: Homo sapiens

US-09-912-559-4

Query Match 100.0%; Score 3153; DB 9; Length 560;

Best Local Similarity 100.0%; Pred. No. 6.5e-240;

Matches 560; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: May 24, 2004, 09:58:58  
Job time : 43 secs